



PCT/GB 00 / 03265



GB 00/03265

INVESTOR IN PEOPLE	
REC'D 27 SEP 2000	
The Patent Office	
WIPO	PCT
Concept House	

Cardiff Road  
Newport  
South Wales  
NP10 8QQ

**PRIORITY DOCUMENT**  
SUBMITTED OR TRANSMITTED IN  
COMPLIANCE WITH  
RULE 17.1(a) OR (b)

I, the undersigned, being an officer duly authorised in accordance with Section 74(1) and (4) of the Deregulation & Contracting Out Act 1994, to sign and issue certificates on behalf of the Comptroller-General, hereby certify that annexed hereto is a true copy of the documents as originally filed in connection with the patent application identified therein.

In accordance with the Patents (Companies Re-registration) Rules 1982, if a company named in this certificate and any accompanying documents has re-registered under the Companies Act 1980 with the same name as that with which it was registered immediately before re-registration save for the substitution as, or inclusion as, the last part of the name of the words "public limited company" or their equivalents in Welsh, references to the name of the company in this certificate and any accompanying documents shall be treated as references to the name with which it is so re-registered.

In accordance with the rules, the words "public limited company" may be replaced by p.l.c., plc, P.L.C. or PLC.

Re-registration under the Companies Act does not constitute a new legal entity but merely subjects the company to certain additional company law rules.

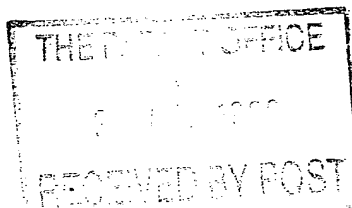
Signed *Ansower.*

Dated 12 September 2000



# Request for grant of a patent

(See the notes on the back of this form. You can also get an explanatory leaflet from the Patent Office to help you fill in this form)



The Patent Office

Cardiff Road  
Newport  
Gwent NP9 1RH

1. Your reference IMPW / P20952GB

2. Patent ap **9920000.8** 25 AUG 1999  
(The Patent number)

3. Full name, address and postcode of the or of each applicant (underline all surnames) Imperial Cancer Research  
Technology  
Sardinia House  
Sardinia Street  
London, WC2A 3NL  
United Kingdom  
Patents ADP number (if you know it) 4113619001

If the applicant is a corporate body, give the country/state of its incorporation United Kingdom 25AUG99 E471890-3 D02866  
P01/7700 0.00 - 9920000.8

4. Title of the invention POLYPEPTIDES

5. Name of your agent (if you have one) ERIC POTTER CLARKSON  
PARK VIEW HOUSE  
58 THE ROPEWALK  
NOTTINGHAM  
NG1 5DD  
"Address for service" in the United Kingdom to which all correspondence should be sent (including the postcode)

Patents ADP number (if you know it) 1305010 ✓

6. If you are declaring priority from one or more earlier patent applications, give the country and the date of filing of the or of each of these earlier applications and (if you know it) the or each application number

Country	Priority application number (if you know it)	Date of filing (day / month / year)

7. If this application is divided or otherwise derived from an earlier UK application, give the number and the filing date of the earlier application

Number of earlier application	Date of filing (day / month / year)

8. Is a statement of inventorship and of right to grant of a patent required in support of this request? (Answer 'Yes' if: YES

- a) any applicant named in part 3 is not an inventor; or
- b) there is an inventor who is not named as an applicant, or
- c) any named applicant is a corporate body. See note (d))

# Patents Form 1/77

9. Enter the number of sheets for any of the following items you are filing with this form. Do not count copies of the same document

Continuation sheets of this form

Description 102

Claims(s) 8

Abstract 1

Drawing(s) 28

10. If you are also filing in any of the following, state how many against each item.

Priority Documents 0

Translations of priority documents 0

Statement of inventorship and right to grant of a patent (*Patents Form 7/77*) NO

Request for preliminary examination and search (*Patents Form 9/77*) NO

Request for substantive examination (*Patents Form 10/77*) NO

Any other documents  
(please specify)

11. I/We request the grant of a patent on the basis of this application.

*Eric Potter Clarkson*

Signature  
ERIC POTTER CLARKSON

Date  
24 August 1999

12. Name and daytime telephone number of person to contact in the United Kingdom 0115 9552211

## Warning

*After an application for a patent has been filed, the Comptroller of the Patent Office will consider whether publication or communication of the invention should be prohibited or restricted under Section 22 of the Patents Act 1977. You will be informed if it is necessary to prohibit or restrict your invention in this way. Furthermore, if you live in the United Kingdom, Section 23 of the Patents Act 1977 stops you from applying for a patent abroad without first getting written permission from the Patent Office unless an application has been filed at least 6 weeks beforehand in the United Kingdom for a patent for the same invention and either no direction prohibiting publication or communication has been given, or any such direction has been revoked.*

## Notes

- If you need help to fill in this form or you have any questions, please contact the Patent Office on 01645 500505.*
- Write your answers in capital letters using black ink or you may type them.*
- If there is not enough space for all the relevant details on any part of this form, please continue on a separate sheet of paper and write "see continuation sheet" in the relevant part(s). Any continuation sheet should be attached to this form.*
- If you have answered 'Yes' Patents Form 7/77 will need to be filed.*
- Once you have filled in the form you must remember to sign and date it.*
- For details of the fee and ways to pay please contact the Patent Office.*

1  
POLYPEPTIDES

The present invention relates to polypeptides and polynucleotides and their use in medicine and screening methods.

5

Members of the TGF- $\beta$  superfamily are secreted signalling molecules produced by cells to influence the behaviour of their neighbours, by regulating cell proliferation, survival, adhesion, differentiation and specification of developmental fate (Hogan *et al* 1994; Kingsley, 1994; Massagué, 1998). These  
10 ligands bind a type II receptor which allows transphosphorylation of the type I receptor (Massagué, 1998). This in turn leads to phosphorylation and activation of the receptor-activated class of Smads (R-Smads; Massagué, 1998), which are responsible for transducing signals from the activated receptors to the nucleus. Smad proteins are a family of highly conserved, intracellular proteins that signal  
15 cellular responses downstream of transforming growth factor-beta (TGF-beta) family serine/threonine kinase receptors. R-Smads 2 and 3 are phosphorylated by TGF- $\beta$  or activin type I receptors, whilst R-Smads 1, 5 and 8 are substrates for BMP type I receptors (Massagué, 1998). Phosphorylation relieves an auto-inhibitory interaction of the C-terminal MH2 domain with the N-terminal MH1  
20 domain (Hata *et al* 1997), allowing the R-Smads to form heteromeric complexes via their MH2 domains with members of the Smad4 class (Lagna *et al* 1996; Zhang *et al* 1997; Masuyama *et al* 1999; Howell *et al* 1999). These activated complexes translocate to the nucleus to regulate transcription of target genes (Whitman, 1998). Smads bind DNA very weakly alone (Shi *et al* 1998) and are  
25 primarily recruited to DNA by other DNA-binding transcription factors (Derynck *et al* 1998; Whitman, 1998), the prototype being the winged-helix/forkhead transcription factor, Fast-1 (Chen *et al* 1996; Chen *et al* 1997).

These co-operating transcription factors are likely to be key determinants of cell type specificity of TGF- $\beta$  signaling, but are mostly still poorly characterized.

The *Xenopus* embryo provides an excellent system in which to elucidate the basis of specificity in TGF- $\beta$  signaling pathways. In the *Xenopus* embryo, TGF- $\beta$  family members act as morphogens, playing key roles in the patterning of different tissues (Green and Smith, 1990; Gurdon *et al* 1994; Hogan, 1996; Whitman, 1998). For example, an activin-like signal, which requires the maternal transcription factor VegT for its production, is released by the vegetal hemisphere of the embryos to induce mesoderm in the overlying equatorial cells (Harland and Gerhart, 1997; Kimelman and Griffin, 1998; Zhang *et al* 1998). The same signaling molecule is also thought to be responsible for specifying endoderm (Henry *et al* 1996). Patterning of the mesoderm and endoderm depends on the precise transcriptional responses of cells within the prospective meso-endoderm to this signal. But what determines which genes are induced in response to this activin-like signal in particular cells, and how is their expression maintained? The presence of particular transcription factors that cooperate with Smads in some cells, but not others could obviously play an important role, as could the presence of other cooperating signaling pathways such as Wnt, FGF and BMP (reviewed by Harland and Gerhart, 1997; Heasman, 1997; Whitman, 1998). The existence in *Xenopus* embryos of multiple transcription factors which are capable of recruiting activin-activated Smads and have different DNA-binding specificity has been proposed, based on the fact that the activin-responsive elements defined in the promoters of differentially expressed meso-endodermal genes share little sequence similarity (reviewed in Howell and Hill, 1997).

The mechanism that confines expression of the *Xenopus goosecoid* gene (Blumberg *et al* 1991) to the dorsal marginal zone of the early gastrula embryo is beginning to be understood. It results from a synergistic interaction between a Wnt signal acting through a proximal element (PE) in the promoter (Watabe *et al* 1995;

- 5 Laurent *et al* 1997) and an activin-like signal acting through a distal element (DE). The DE is also conserved in the mouse and zebrafish *goosecoid* promoters (Watabe *et al* 1995; Candia *et al* 1997; McKendry *et al* 1998). Since the sequence of the DE bears no resemblance to the ARE from the *Mix.2* promoter, the transcription factors involved in its activin-inducibility may be distinct from Fast-
- 10 1. A paired-like homeodomain factor of unknown identity has been implicated in the activin-responsive transcription of the DE-related element in the zebrafish *goosecoid* promoter (McKendry *et al* 1998).

The TGF $\beta$  superfamily, signalling pathways and likely functions have been

15 extensively researched and reviewed. TGF $\beta$  appears to be involved in the modulation of many biological processes and may be implicated in pathogenic conditions including tumour growth, inflammation, wound healing, scarring, fibrosis, kidney damage, for example in diabetes, and atherosclerosis. Proteins related to TGF $\beta$  include activins, inhibins and bone morphogenetic proteins

20 (BMPs). In some situations, enhancement of TGF $\beta$  signalling may be beneficial, whilst in others, inhibition may be useful. A lack of specific small-molecule agonists or antagonists of TGF $\beta$  signalling has impeded investigations, particularly *in vivo*.

25 The views expressed in a selection of reviews are summarised below.

Hartsough MT; Mulder KM (1997) "Transforming growth factor- $\beta$  signalling in epithelial cells" *Pharmacol Ther* **75** (1), 21-41 discusses the resistance of some tumours to growth suppression by TGF $\beta$ .

- 5 Noble NA; Border WA (1997) "Angiotensin II in renal fibrosis: should TGF- $\beta$  rather than blood pressure be the therapeutic target?" *Semin Nephrol* **17**(5), 455-66 discusses the role of TGF $\beta$  in promoting tissue fibrosis and the induction of TGF $\beta$  by angiotensin II.
- 10 Koli K; Keski-Oja J (1996) "Transforming growth factor- $\beta$  system and its regulation by members of the steroid-thyroid hormone superfamily." *Adv Cancer Res* **70**, 63-94, discusses TGF- $\beta$ s and their receptors and their action as key regulators of many aspects of cell growth, differentiation, and function, particularly malignancy.
- 15 Grande JP (1997) "Role of transforming growth factor- $\beta$  in tissue injury and repair." *Proc Soc Exp Biol Med* **214**(1), 27-40 discusses the role of TGF $\beta$  in normal cell growth, development, and tissue remodelling following injury. Disruption of the TGF $\beta$ 1 gene *in utero* produces a wasting syndrome
- 20 characterised by systemic inflammation, suggesting that this growth factor plays an important role in limiting the inflammatory response. TGF $\beta$  is a dominant mediator of the pathologic extracellular matrix accumulation that characterises progression of tissue injury to end-stage organ failure. Recent studies directed towards characterisation of the TGF $\beta$  genes, dissection of the mechanisms by
- 25 which TGF $\beta$ s are produced and activated, and identification of TGF $\beta$  signalling pathways have established the important roles that these family members play in



cell and tissue homeostasis. TGF $\beta$  structure-function relationships and their relevance to models of tissue injury/wound repair are also discussed.

Lawrence DA (1996) "Transforming growth factor- $\beta$ : a general review." *Eur Cytokine Netw* 7(3), 363-74 reviews the roles of TGF- $\beta$ 1,  $\beta$ 2 and  $\beta$ 3 in mammals. The author comments that they play critical roles in growth regulation and development. All three of these growth factors are secreted by most cell types, generally in a latent form, requiring activation before they can exert biological activity. This activation of latent TGF- $\beta$ , which may involve plasmin, thrombospondin and possibly acidic microenvironments, appears to be a crucial regulatory step in controlling their effects. The TGF- $\beta$ s possess three major activities: they inhibit proliferation of most cells, but can stimulate the growth of some mesenchymal cells; they exert immunosuppressive effects; and they enhance the formation of extracellular matrix. Two types of membrane receptors (type I and type II) possessing a serine/threonine kinase activity within their cytoplasmic domains are involved in signal transduction. Inhibition of growth by the TGF- $\beta$ s stems from a blockage of the cell cycle in late G1 phase. Among the molecular participants concerned in G1-arrest are the Retinoblastoma (Rb) protein and members of the Cyclin/Cyclin-dependent kinase/Cyclin dependent kinase inhibitor families. In the intact organism the TGF- $\beta$ s are involved in wound repair processes and in starting inflammatory reactions and then in their resolution. The latter effects of the TGF- $\beta$ s derive in part from their chemotactic attraction of inflammatory cells and of fibroblasts. From gene knockout and from overexpression studies it has been shown that precise regulation of each isoform is essential for survival, at least in the long term. Several clinical applications for certain isoforms have already shown their efficacy and they have been implicated in numerous other pathological situations.

Pignatelli M; Gilligan CJ (1996) "Transforming growth factor- $\beta$  in GI neoplasia, wound healing and immune response." *Baillieres Clin Gastroenterol* **10(1)**, 65-81 discusses the influence that cell-cell and cell-matrix interactions, the differentiating status of the cell together with the functional activity of other soluble growth factors have on responses to TGF- $\beta$ s, particularly in relation to homeostasis of the GI mucosa and their role in gastrointestinal carcinogenesis.

Cox DA (1995) "Transforming growth factor- $\beta$  3." *Cell Biol Int* **19(5)**, 357-71 discusses the molecular and cellular biology of TGF- $\beta$  3 and those physiological actions which may lead to clinical applications, particularly in the indication areas of wound healing and chemoprotection.

Wahl SM (1992) "Transforming growth factor  $\beta$  (TGF- $\beta$ ) in inflammation: a cause and a cure." *J Clin Immunol* **12(2)**, 61-74 discusses the mechanisms controlling whether the pro- or antiinflammatory effects of this peptide prevail.

Ruscetti FW; Palladino MA (1991) "Transforming growth factor- $\beta$  and the immune system." *Prog Growth Factor Res* **3(2)**, 159-75 discusses the increased levels of TGF- $\beta$  found in several disease states associated with immunosuppression such as different forms of malignancy, chronic degenerative diseases, and AIDS, implicating the involvement of TGF- $\beta$  in the pathogenesis of some diseases.

TGF $\beta$  is known to be an inhibitor of inflammation (as reviewed, for example, in Lawrence (1996) and Grande (1997), both cited above) for example from studies

in which massive inflammatory lesions are seen in mice in which a TGF $\beta$  gene is inactivated.

Here we identify new partners for activated Smads. We have identified a short  
5 motif, characterized by the sequence PP(T/N)K, that is necessary and may be  
sufficient for interaction with the MH2 domain of Smad2. Full-length Smad  
polypeptides, for example Smad2 and Smad3, may be activated by  
phosphorylation near the C-terminus of the polypeptide, which induces a  
conformational change which exposes a binding site in the MH2 domain for  
10 transcription factors such as FAST1, FAST2 or the newly-identified partners. A  
Smad polypeptide in which the N-terminal domain is not present or is truncated  
may not require phosphorylation in order to expose this binding site.

A first aspect of the invention provides a polypeptide (interacting polypeptide)  
15 capable of interacting with a Smad polypeptide wherein the interacting  
polypeptide comprises the amino acid sequence PP(T/N)K and is less than 32,  
31, or 30 amino acids in length. The interacting polypeptide may alternatively  
comprise the amino acid sequence PPSK or PPQK ie a residue with an aliphatic  
hydroxyl side chain or an amide side chain may be present between the PP and K  
20 residues. By "interacting with" is included the meaning of "binding to", for  
example detectably binding to, for example binding detectable using any method  
of detecting protein/protein binding as indicated below, for example co-  
immunoprecipitation or a surface plasmon resonance technique. The term  
"polypeptide" in connection with the interacting polypeptide includes peptides as  
25 small as the peptide PPNK or PPTK. The invention includes a polypeptide of  
less than 32, 31 or 30 amino acids in length comprising the amino acid sequence  
PP(T/N)K.

A further aspect of the invention provides a polypeptide (interacting polypeptide) capable of interacting with a Smad polypeptide wherein the interacting polypeptide comprises the amino acid sequence PP(T/N)K and is not full-length  
 5 *Xenopus* or human FAST1 or a fragment thereof, mouse FAST2, *Xenopus* Milk, *Xenopus* Mixer, *Xenopus* Bix3, Bix2 or Bix1. The interacting polypeptide may be *Xenopus* FAST3, the sequence of which is shown in Figure 13.

The terms FAST1, FAST2, Milk, Mixer, Bix3, Bix2 and Bix1 are well known  
 10 to those skilled in the art. Mixer may also be known as Mix3. The sequence for FAST2 is given, for example, in Liu *et al* (1999) *Mol Cell Biol* 19, 424-430 or Labbé *et al* (1998) *Mol Cell* 2, 109-120. The sequence for FAST1 is given, for example, in Chen *et al* (1996) *Nature* 383, 691-696 and Chen *et al* (1997) *Nature* 389, 85-89. The sequence of human Fast1 is given in Zhou *et al* (1998)  
 15 *Mol Cell* 2, 121-127 and in WO98/5380. Fragments of FAST1 are described in Chen *et al* (1997) *Nature* 389, 85-89 and in WO98/5380. The sequence for Milk is given in Ecochard *et al* (1998). The sequence for Mixer is given in Henry & Melton (1998). The sequences for Bix3, Bix2 and Bix1 are given in Tada *et al* (1998). Bix1 may also be known as Mix4 (see Mead *et al* (1998) Cloning of  
 20 Mix-related homeodomain proteins using fast retrieval of gel shift activities, (FROGS), a technique for the isolation of DNA-binding proteins *Proc Natl Acad Sci U S A* 95(19), 11251-6). The sequences of these polypeptide are also shown in Figure 13.

25 As discussed further below, the residue immediately before (ie N-terminal of) the amino acid sequence PP(T/N)K may preferably be a hydrophobic residue, for example F, M or V. The residue immediately after (ie C-terminal of) the

amino acid sequence PP(T/N)K ie at position +1 may preferably be an S or T, which may be immediately followed by an I or V residue. An acidic residue (for example glutamate or aspartate) may be present at position about +3 to about +10, preferably +4 or +5 and may be immediately followed by a hydrophobic residue, for example M, V or I.

The interacting polypeptide may be a transcription factor or a fragment thereof. Thus, the interacting polypeptide may comprise a domain that is capable of binding to a nucleic acid, preferably DNA, still more preferably double-stranded DNA, yet more preferably to DNA that forms part of a promoter region for a gene. The interacting polypeptide may be a fragment of a transcription factor wherein the transcription factor comprises a said domain that is capable of binding to a nucleic acid but the interacting polypeptide does not comprise the said domain. It will be appreciated that the interacting polypeptide may bind to the said nucleic acid with higher affinity when the interacting polypeptide is bound to one or more other polypeptides, for example one or more Smad polypeptides, than when it is not so bound. The interacting polypeptide may bind to the said nucleic acid as a dimer or as a heterodimer with another transcription factor ie with another polypeptide comprising a domain that is capable of binding to a nucleic acid. The interacting polypeptide may be capable of promoting transcription of DNA; additional polypeptides may be required for transcription to take place. The interacting polypeptide may comprise, for example, a winged-helix DNA binding domain or a Paired DNA binding domain or a homeodomain, for example a Paired-like homeodomain. It will be appreciated that the interacting polypeptide may comprise more than one domain that is capable of binding to a nucleic acid.

As is well known to those skilled in the art, a promoter is an expression control element formed by a DNA sequence that permits binding of RNA polymerase and transcription to occur. A promoter may be a region of DNA capable of controlling transcription of neighbouring DNA. It will be appreciated that a transcription factor that is capable of interacting with a Smad polypeptide may not be capable of binding to DNA unless it is in a complex with the Smad polypeptide, for example Smad2 or Smad3 and Smad4. The transcription factor may be capable of interacting (directly or indirectly) with an RNA polymerase. It is preferred that the transcription factor is capable of interacting directly with an RNA polymerase.

FAST1 and FAST2 comprise a winged-helix (also known as a Forkhead) DNA binding domain. Members of the Mix family, which may include the chicken CMIX polypeptide (Peale *et al* (1998) *Mech of Dev* **75**, 167-170 and Stein *et al* (1998) *Mech of Dev* **75**, 163-165), comprise a Paired-like homeodomain (see, for example Wilson *et al* (1993) *Genes Dev* **7**, 2120-2134).

The term paired homeodomain transcription factor is well known to those skilled in the art. Paired homeodomain transcription factors are reviewed in, for example, Galliot *et al* (1999) Evolution of homeobox genes: Q50 Paired-like genes founded the Paired class *Dev Genes Evol* **209**, 186-197, Wright *et al* (1989) Vertebrate homeodomain proteins: families of region-specific transcription factors *Trends Biochem Sci* **14**, 52-56 and Dorn *et al* (1994) Homeodomain proteins in development and therapy *Pharmacol Ther* **61**, 155-184.

The homeobox domain has about 60 amino acids and consists of a helix-turn-helix motif that binds DNA by inserting the recognition helix into the major groove of the DNA and its amino-terminal arm into the adjacent minor groove. Representative homeobox domain are found in the *Drosophila* Antennapedia polypeptide and the *Drosophila* Paired polypeptide.

Galliot *et al* (1999) *Dev Genes Evol* **209**, 186-197 reviews polypeptides belonging to the Paired class. This class of polypeptides contain a homeobox DNA binding domain that is related to that found in the *Drosophila* gene Paired (*prd*) and characterised by invariant residues which distinguish them from other homeodomain (HD) classes. Three subclasses can be defined according to the residue at position 50 of the homeodomain, which plays a key role in determining DNA binding specificity. The Pax or Prd-type genes have a serine residue at position 50 (S<sub>50</sub> type) and also have a second DNA-binding domain, the *prd* (Paired) domain. Mammalian members of this sub-class include the Pax genes (see, for example, Adams *et al* (1992) *Genes & Dev* **6**, 1589-1607). A second sub-class has a lysine at position 50 (K<sub>50</sub> type) and a third sub-class has a glutamine residue (Q<sub>50</sub> type) at position 50. The K<sub>50</sub> and Q<sub>50</sub> sub-classes do not have the *prd* domain. The Mix family of polypeptides belongs to the Q<sub>50</sub> class.

The paired domain motif is a domain of 128 amino acids identified as a secondary homology region in the homeobox-containing proteins of the *Drosophila* *paired* and *gooseberry* genes (Bopp *et al* (1986) *Cell* **47**, 1033-1040; Baumgartner *et al* (1987) *Genes & Dev* **1**, 1247-1267). The paired domain motif encodes a DNA-binding motif (Goulding *et al* (1991) *EMBO J* **10**, 1135-1147; Treisman *et al* (1991) *Genes & Dev* **5**, 594-604; Chalepakidis *et al* (1991) *Cell* **66**, 873-884). Three  $\alpha$ -helices are predicted to be present in the paired domain (see

Bopp *et al* (1989) *EMBO J* 8, 3447-3457). The paired domain proteins of vertebrates are encoded by a multigene family that has been conserved in evolution, termed the Pax gene family, as mentioned above.

- 5 The term Forkhead or winged helix polypeptide is well known to those skilled in the art. Forkhead/winged helix polypeptides are reviewed, for example, in Kaufmann & Knochel (1996) *Mech Dev* 57, 3-20. A polypeptide may be identified as a Forkhead or winged-helix polypeptide if it comprises a domain with features of a Forkhead/winged-helix DNA binding domain. The
- 10 Forkhead/winged-helix domain is a variant of the helix-turn-helix motif (Brennan (1993) The winged-helix DNA-binding motif: Another helix-turn-helix takeoff *Cell* 74, 773-776; Clark *et al* (1993) Co-crystal structure of the HNF-3/forkhead DNA-recognition motif resembles histone H5 *Nature* 364, 412-420). The forkhead/winged-helix domain is responsible for DNA-binding specificity
- 15 and binds to DNA as a monomer, with two loops or wings on the C-terminal side of the helix-turn-helix.

- The forkhead domain is about 111 amino acids in length. Based on the degree of homology within the forkhead domain, the forkhead family is further split into
- 20 subgroups. Over 80 genes with the conserved wing-helix forkhead motif have been identified from yeast to mammalian sources, as reviewed in Kaufmann & Knochel (1996) *Mech Dev* 57, 3-20. Sequence identity in the 111 amino acid domain may be more than about 50%, for example between about 70% and 95% identity; sequence identity outside this domain between forkhead family members
- 25 may be much less. A Forkhead protein may have at least 30, 40, 50, 60, 75, 80, 85, 90 or 95% amino acid sequence identity with the FKHR Forkhead domain (Davis *et al* (1995) *Hum Mol Genet* 4, 2355-2362).



The Forkhead domains of FAST1 and FAST2 are about 40% identical to that of HNF-3 $\beta$  and several other family members (Liu *et al* (1999). FAST1 and FAST2 are highly homologous in the Forkhead domain and have sequence  
 5 similarity in other domains. No homology to other Forkhead proteins is observed outside the Forkhead domain. FAST polypeptides therefore appear to form a sub-family of the Forkhead family.

The partial sequence of a novel FAST polypeptide, termed *Xenopus* FAST 3 is  
 10 shown in Figure 13.

It will be appreciated that the interacting polypeptide may bind Smad2 and/or Smad3 MH2 domains but may not bind Smad1 or Smad4 directly. The interaction may require the  $\alpha$ -helix 2 of the MH2 domain, though the interaction  
 15 may not be with the  $\alpha$ -helix 2. The interaction may require regions equivalent to the regions of Smad2 indicated in Table 1 to be required for the interactions investigated.

It is preferred that the Smad polypeptide with which the interacting polypeptide  
 20 interacts is Smad2 or Smad3, more preferably human Smad2 or human Smad3, most preferably human Smad2. The terms Smad, Smad2 and Smad3 are well known to those skilled in the art; see, for example Massagué (1998); Macias-Silva *et al* (1996) *Cell* **87**, 1215-1224 (human Smad2); Graff *et al* (1996) *Cell* **85**, 479-487 (*Xenopus* Smad2); Zhang *et al* (1996) *Nature* **383**, 168-172 (human Smad3).  
 25 The sequence of *Xenopus* Smad3, a novel Smad polypeptide, is shown in Figure 12 with the sequences of human Smads 2 and 3 and *Xenopus* Smad2.

It will be appreciated that a Smad polypeptide may have a domain recognisable as an MH2 domain. The MH2 domains of *Drosophila*, *Xenopus*, human and mouse Smad2, for example, appear to be more than 90% identical (Brummel *et al* (1999) *Genes Dev* 13, 98-111). A tryptophan residue may be present at the residue equivalent to W274 of *Xenopus* Smad2 (see, for example, WO97/22697). Smads 1, 2, 3, 4, 5 and 8 may further have a conserved domain recognisable as a MH1 domain, whilst Smads 6 and 7 may have a divergent MH1 domain. Smads 2 and 3 may be activated by TGF $\beta$  or activin by phosphorylation at two serine residues near the C-terminus of the polypeptide. Smads 2 and 3 may be cytoplasmic until activated and then translocate to the nucleus. Smads 2 and 3 may also form a complex with Smad4 in response to ligand.

In terms of sequence, Smad2 and 3 may be defined by the sequence in the L3 loop which may dictate their binding to the activin and TGF $\beta$  type I receptors and the sequence of the  $\alpha$ -helix 2 that is required to bind to Fast1 (see Shi *et al* (1997) *Nature* 388, 87-93 and WO99/01765), Milk and Mixer (see below).

The Smad polypeptide may be a variant, fragment, derivative or fusion of human Smad2 or human Smad3.

It is preferred that the Smad polypeptide has a greater amino acid identity with the C-terminal MH2 region, particularly the  $\alpha$ -helix2 region (see Chen *et al* (1998)), of Smad2 or Smad3, for example human Smad2 or Smad3, than with the C-terminal MH2 region, particularly the  $\alpha$ -helix2 region, of Smad1 or Smad4, for example human Smad1 or human Smad4. The MH2 domain of *Xenopus* Smad2 starts at amino acid W-274.

By “variants” of a polypeptide, for example of Smad2 or Smad3, we include insertions, deletions and substitutions, either conservative or non-conservative. In particular we include variants of the polypeptide where such changes do not substantially alter the activity of the said polypeptide, for example the ability of the Smad polypeptide to bind to an interacting polypeptide, for example a transcription factor such as FAST1, FAST2, Mixer or Milk, or another Smad polypeptide, for example Smad4.

By “conservative substitutions” is intended combinations such as Gly, Ala; Val, Ile, Leu; Asp, Glu; Asn, Gln; Ser, Thr; Lys, Arg; and Phe, Tyr.

It is particularly preferred if the Smad polypeptide variant has an amino acid sequence which has at least 65% identity with the amino acid sequence of Smad2 or Smad3, for example the amino acid sequence of Smad2 or Smad3 shown in Macias-Silva *et al* (1996) *Cell* **87**, 1215-1224 (human Smad2); Graff *et al* (1996) *Cell* **85**, 479-487 (*Xenopus* Smad2); Zhang *et al* (1996) *Nature* **383**, 168-172 (human Smad3) or Figure 12 (*Xenopus* Smad3), more preferably at least 50%, 55%, 60%, 70%, still more preferably at least 75%, yet still more preferably at least 80%, in further preference at least 85%, in still further preference at least 90% and most preferably at least 95% or 97% identity with the amino acid sequence defined above.

It is still further preferred if the Smad polypeptide variant has an amino acid sequence which has at least 65% identity with the amino acid sequence of the  $\alpha$ -helix2 domain of Smad2 or Smad3 shown in Macias-Silva *et al* (1996) *Cell* **87**, 1215-1224 (human Smad2); Graff *et al* (1996) *Cell* **85**, 479-487 (*Xenopus* Smad2); Zhang *et al* (1996) *Nature* **383**, 168-172 (human Smad3) or Figure 12

(*Xenopus* Smad3), more preferably at least 70% or 73%, still more preferably at least 75%, yet still more preferably at least 80%, in further preference at least 83% or 85%, in still further preference at least 90% and most preferably at least 95% or 97% identity with the amino acid sequence defined above. It will be appreciated that the  $\alpha$ -helix2 domain of a Smad polypeptide may be readily identified by a person skilled in the art and as described in Chen *et al* (1998), for example using sequence comparisons as described below.

The percent sequence identity between two polypeptides may be determined using suitable computer programs, for example the GAP program of the University of Wisconsin Genetic Computing Group and it will be appreciated that percent identity is calculated in relation to polypeptides whose sequence has been aligned optimally.

The alignment may alternatively be carried out using the Clustal W program (Thompson *et al* (1994) *Nucl Acid Res* 22, 4673-4680). The parameters used may be as follows:

Fast pairwise alignment parameters: K-tuple(word) size; 1, window size; 5, gap penalty; 3, number of top diagonals; 5. Scoring method: x percent.

Multiple alignment parameters: gap open penalty; 10, gap extension penalty; 0.05.

Scoring matrix: BLOSUM.

"Variations" of the polypeptide also include a polypeptide in which relatively short stretches (for example 5 to 20 amino acids) have a high degree of homology (at least 80% and preferably at least 90 or 95%) with equivalent stretches of the polypeptide even though the overall homology between the two

polypeptides may be much less. This is because important active or binding sites may be shared even when the general architecture of the protein is different.

It is preferred that the Smad polypeptide, for example Smad2 or Smad3  
5 polypeptide is a polypeptide which consists of the amino acid sequence of the Smad2 or Smad3 polypeptide as shown in Macias-Silva *et al* (1996) *Cell* **87**, 1215-1224 (human Smad2); Graff *et al* (1996) *Cell* **85**, 479-487 (*Xenopus* Smad2); Zhang *et al* (1996) *Nature* **383**, 168-172 (human Smad3) or Figure 12 (*Xenopus* Smad3), or naturally occurring allelic variants thereof and fusions  
10 thereof. A preferred fusion may be a GST fusion, for example as described in Example 1 or any other fusion described in Example 1 or a Myc fusion as described, for example, in Chen *et al* (1997). A further preferred fusion may have the tag Glu-Phe-Met-Pro-Met-Glu (termed EE-tag) or a His, HA or FLAG tag, as well known to those skilled in the art.

15 Alternatively, it is preferred that the Smad polypeptide is a fragment or a fusion of a fragment of a Smad2 or Smad3 polypeptide, as shown in Macias-Silva *et al* (1996) (human Smad2); Graff *et al* (1996) (*Xenopus* Smad2); Zhang *et al* (1996) (human Smad3) or Figure 12 (*Xenopus* Smad3), or naturally occurring allelic  
20 variants thereof. It is preferred that the said fragment or fusion of a fragment comprises the MH2 domain, in particular the  $\alpha$ -helix 2 domain of the said Smad2 or Smad3 polypeptide, as shown in the references indicated above, or naturally occurring allelic variants thereof. Particularly preferred fragments or fusions include the fragments indicated in Table 1 as capable of binding to the  
25 endogenous activity, Mixer, Milk or Fast-1, and fusions of those fragments, for example with GST.

It is preferred that the Smad polypeptide is a polypeptide that is capable of binding to FAST1, FAST2, FAST3, Mixer, Milk, Bix1 or Bix3. The capability of the said Smad polypeptide with regard to binding FAST1, FAST2, FAST3, Mixer, Milk, Bix1 or Bix3 may be measured by any method of detecting/measuring a protein/protein interaction, as discussed further below and in Example 1. Suitable methods include yeast two-hybrid interactions, co-purification (for example co-immunoprecipitation or GST-pulldown assays), ELISA, co-immunoprecipitation methods and bandshift assays.

It will be appreciated that it may be necessary for the Smad polypeptide to be phosphorylated in order for FAST1, FAST2, Mixer, Milk, Bix1 or Bix3 or the said interacting polypeptide, for example FAST3, to be capable of binding to the Smad polypeptide ie for the Smad polypeptide to be activated. Phosphorylation of a full-length Smad polypeptide may be necessary to relieve an auto-inhibitory interaction of the C-terminal MH2 domain with the N-terminal MH1 domain, as discussed above. Smad fragments in which the N-terminal MH1 domain is absent, disrupted or truncated may not require phosphorylation in order for the interacting polypeptide to interact with the fragment. The relevant phosphorylation of Smad2 takes place on residues Ser465 and Ser467 (see, for example, Souchelnyskyi *et al* (1997) *J Biol Chem* 272, 28107-28115). Phosphorylation may be performed *in vitro*, for example by immunoprecipitating active receptor complexes from Cos1 cells overexpressing the receptors and treated with TGF $\beta$ . These immunoprecipitates will phosphorylate GST-Smad2, for example as described in Macias-Silva *et al* (1996) *Cell* 87, 1215-1224. It is preferred that the Smad polypeptide is a polypeptide, for example a fragment in which the N-terminal MH1 domain is absent, disrupted or truncated, that does not require phosphorylation in order to be able to bind to FAST1, FAST2,

Mixer, Milk, Bix 1, 2 or 3 or the said interacting polypeptide, for example FAST3. Suitable Smad polypeptides may be the preferred fragments and fusions capable of binding to the endogenous activity, Mixer, Milk or FAST1 listed in Table 1.

5

The interacting polypeptide may be capable of interacting with a portion of the Smad polypeptide that is equivalent to  $\alpha$ -helix 2 or part thereof of a full length Smad polypeptide, for example Smad2 or Smad3.

- 10 It is preferred that the interacting polypeptide or PP(T/N)K- containing polypeptide is less (in order of preference) than 150, 100, 80, 70, 50, 40, 30 or 26 amino acids in length. It is further preferred that the interacting polypeptide is at least (in order to preference) 4, 5, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24 or 25 amino acids in length or any combination of these maximum and minimum
- 15 lengths. It is particularly preferred that the interacting polypeptide is between 4 and about 30 amino acids in length; in further preference the interacting polypeptide is between 25 and about 30 amino acids in length.

- It is preferred if the interacting polypeptide consists of a fragment of a naturally occurring protein such as those described below or a fusion thereof. Suitably, the fragment of a naturally occurring protein is less than (in order of preference) 150, 100, 80, 70, 50, 40, 30 or 26 amino acids in length. Also suitably, the fragment of a naturally occurring protein is at least (in order of preference) 4, 5, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24 or 25 amino acids in length. It is preferred
- 25 that the interacting polypeptide further has an acidic (ie negatively charged) amino acid residue present at a position from 3 to 10, preferably 4 to 5 residues C-terminal of the amino acid sequence PP(T/N)K and/or a proline residue

present at a position from 5 to 20 residues C-terminal of the amino acid sequence PP(T/N)K. The acidic (negatively charged) amino acid residue is typically a glutamate or aspartate residue. The two proline residues within the PP(T/N)K motif are believed to be essential for interaction with the Smad polypeptide.

- 5 Polypeptides with the sequences AANK or QTNK in place of PP(T/N)K appear not to bind Smad2. The downstream proline and aspartate residues as described above may also be important for binding. As discussed above, the residue immediately before (ie N-terminal of) the amino acid sequence PP(T/N)K may preferably be a hydrophobic residue, for example F, M or V. The residue
- 10 immediately after (ie C-terminal of) the amino acid sequence PP(T/N)K ie at position +1 may preferably be an S or T, which may be immediately followed by an I or V residue. An acidic residue (for example glutamate or aspartate) may be present at position about +3 to about +10, preferably +4 or +5, and may be immediately followed by a hydrophobic residue, for example M, V or I.

15

It is particularly preferred that the interacting polypeptide consists of or comprises the amino acid sequence PPNKTITPDMNVRIPPI or PPNKTITPDMNTIIPQI or PPNKSVFDVLTSHPGD or PPNKSIYDVWVSHPRD or PPTKTITANMNTIIPQM or

20 PPNKSIYDVWVSHPRD or PPNKTVFDIPVYTGHGP or PPNKTITPDMNTIIPQI or LLMDFNFPNKTITPDMNVRIPPI or HSNLMMDFPNKTITPDMNTIIPQI or LDNMLRAMPPNKSVDVLTSHPGD or LDSLFQGVPPNKSIVDVWVSHPRD or

25 LMMDISNFPPTKTITANMNTIIPQM or LDALFQGVPPNKSIVDVWVSHPRD or LKNAPSDFPNKTIVFDIPVYTGHGP or



HSNLVMEFPPNKTITPDMNTIIPQI. The interacting polypeptide may consist of or comprise the amino acid sequence of residues 283 to 307 of *Xenopus* Mixer, residues 316 to 340 of *Xenopus* Milk, residues 470 to 493 of *Xenopus* FAST1, residues 363 to 386 of mouse FAST2, residues 316 to 341 of *Xenopus* Bix2, residues 305 to 319 of *Xenopus* Bix 3, residues 327 to 350 of human FAST1, residues 363 to 386 of human FAST1, residues 245 to 269 of *Xenopus* FAST3, residues 319 to 343 of *Xenopus* Bix1 or the equivalent residues of the equivalent mammalian, preferably human, Mixer, Milk, Bix, FAST1, FAST2 or FAST3 polypeptides.

10

The interacting polypeptide or PP(T/N)K-containing polypeptide typically comprises the amino acid sequence  $X_n\text{PP(T/N)K}Z_m$  wherein  $X_n$  represents the amino acid sequence of the consecutive  $n$  amino acids immediately N terminal to the amino acid sequence PP(T/N)K in a naturally occurring polypeptide comprising the amino acid sequence PP(T/N)K, for example a said naturally occurring polypeptide described above, and wherein  $Z_m$  represents the amino acid sequence of the consecutive  $m$  amino acids immediately C terminal to the amino acid sequence PP(T/N)K in a naturally occurring polypeptide comprising the amino acid sequence PP(T/N)K, for example a said naturally occurring polypeptide described above, wherein  $n$  and  $m$  may independently be any number between 0 and 1, 5, 10, 15, 20, 25, 30, 50, 80, 100, 150, 200, 300 or 500 amino acids, preferably between 0 and 150, still more preferably between 0 and 30 amino acids. It is preferred that the amino acid sequences  $X_n$  and  $Z_m$  are immediately N and C terminal, respectively, to the amino acid sequence PP(T/N)K in the same naturally occurring polypeptide.

25

By “residue equivalent to” a particular residue, for example the residue Pro291 of full-length *Xenopus* Mixer, is included the meaning that the amino acid residue occupies a position in the native two or three dimensional structure of a polypeptide, for example a transcription factor comprising a Paired-like homeodomain, corresponding to the position occupied by the said particular residue, for example Pro291, in the native two or three dimensional structure of full-length *Xenopus* Mixer. It will be appreciated that Pro291 of *Xenopus* full-length Mixer is located outside the Paired-like homeodomain, towards the C-terminus of the polypeptide.

The residue equivalent to a particular residue, for example the residue Pro291 of full-length *Xenopus* Mixer, may be identified by alignment of the sequence of the polypeptide with that of full-length *Xenopus* Mixer in such a way as to maximise the match between the sequences. The alignment may be carried out by visual inspection and/or by the use of suitable computer programs, for example the GAP program of the University of Wisconsin Genetic Computing Group, which will also allow the percent identity of the polypeptides to be calculated. The Align program (Pearson (1994) in: Methods in Molecular Biology, Computer Analysis of Sequence Data, Part II (Griffin, AM and Griffin, HG eds) pp 365-389, Humana Press, Clifton). Thus, residues identified in this manner are also “equivalent residues”.

It will be appreciated that in the case of truncated forms of Mixer or in forms where simple replacements of amino acids have occurred it is facile to identify the “equivalent residue”.

The sequence for *Xenopus* Mixer is given in, for example, Henry & Melton (1998).

5 The three-letter and one-letter amino acid code of the IUPAC-IUB Biochemical Nomenclature Commission is used herein. The sequence of polypeptides are given N-terminal to C-terminal as is conventional. In particular, Xaa represents any amino acid. It is preferred that the amino acids are L-amino acids, in particular it is strongly preferred that the PP(T/N)K motif consists of L-amino acid residues. It is preferred that the amino acid residues immediately flanking  
10 (such as those within 10 to 20 residues) of the PP(T/N)K motif are L-amino acids residues, but that they may be D-amino acid residues.

The above polypeptides or peptide may be made by methods well known in the art and as described below and in Example 1, for example using molecular  
15 biology methods or automated chemical peptide synthesis methods.

Peptides may be synthesised by the Fmoc-polyamide mode of solid-phase peptide synthesis as disclosed by Lu *et al* (1981) *J. Org. Chem.* **46**, 3433 and references therein. Temporary N-amino group protection is afforded by the 9-  
20 fluorenylmethyloxycarbonyl (Fmoc) group. Repetitive cleavage of this highly base-labile protecting group is effected using 20% piperidine in N,N-dimethylformamide. Side-chain functionalities may be protected as their butyl  
ethers (in the case of serine threonine and tyrosine), butyl esters (in the case of glutamic acid and aspartic acid), butyloxycarbonyl derivative (in the case of lysine  
25 and histidine), trityl derivative (in the case of cysteine) and 4-methoxy-2,3,6-trimethylbenzenesulphonyl derivative (in the case of arginine). Where glutamine or asparagine are C-terminal residues, use is made of the 4,4'-

dimethoxybenzhydryl group for protection of the side chain amido functionalities. The solid-phase support is based on a polydimethyl-acrylamide polymer constituted from the three monomers dimethylacrylamide (backbone-monomer), bisacryloylethylene diamine (cross linker) and acryloylsarcosine methyl ester (functionalising agent). The peptide-to-resin cleavable linked agent used is the acid-labile 4-hydroxymethyl-phenoxyacetic acid derivative. All amino acid derivatives are added as their preformed symmetrical anhydride derivatives with the exception of asparagine and glutamine, which are added using a reversed N,N-dicyclohexyl-carbodiimide/1-hydroxybenzotriazole mediated coupling procedure.

All coupling and deprotection reactions are monitored using ninhydrin, trinitrobenzene sulphonic acid or isotin test procedures. Upon completion of synthesis, peptides are cleaved from the resin support with concomitant removal of side-chain protecting groups by treatment with 95% trifluoroacetic acid containing a 50% scavenger mix. Scavengers commonly used are ethanedithiol, phenol, anisole and water, the exact choice depending on the constituent amino acids of the peptide being synthesised. Trifluoroacetic acid is removed by evaporation *in vacuo*, with subsequent trituration with diethyl ether affording the crude peptide. Any scavengers present are removed by a simple extraction procedure which on lyophilisation of the aqueous phase affords the crude peptide free of scavengers.

Reagents for peptide synthesis are generally available from Calbiochem-Novabiochem (UK) Ltd, Nottingham NG7 2QJ, UK. Purification may be effected by any one, or a combination of, techniques such as size exclusion chromatography, ion-exchange chromatography and (principally) reverse-phase high performance liquid chromatography. Analysis of peptides may be carried out using thin layer chromatography, reverse-phase high performance liquid chromatography, amino-acid analysis after acid hydrolysis and by fast atom bombardment (FAB) mass spectrometric analysis.

It will be appreciated that peptidomimetic compounds may also be useful. Thus, by "polypeptide" or "peptide" we include not only molecules in which amino acid residues are joined by peptide (-CO-NH-) linkages but also molecules in which the peptide bond is reversed. Such retro-inverso peptidomimetics may be made using methods known in the art, for example such as those described in Mézière *et al* (1997) *J. Immunol.* 159, 3230-3237, incorporated herein by reference. This approach involves making pseudopeptides containing changes involving the backbone, and not the orientation of side chains. Mézière *et al* (1997) show that, at least for MHC class II and T helper cell responses, these pseudopeptides are useful. Retro-inverse peptides, which contain NH-CO bonds instead of CO-NH peptide bonds, are much more resistant to proteolysis.

Similarly, the peptide bond may be dispensed with altogether provided that an appropriate linker moiety which retains the spacing between the C $\alpha$  atoms of the amino acid residues is used; it is particularly preferred if the linker moiety has substantially the same charge distribution and substantially the same planarity as a peptide bond.

It will be appreciated that the peptide may conveniently be blocked at its N- or C-terminus so as to help reduce susceptibility to exoproteolytic digestion.

Thus, it will be appreciated that the interacting polypeptide, for example which comprises the amino acid sequence PP(T/N)K may be a peptidomimetic compound, as described above.

A further aspect of the invention provides a molecule comprising an interacting polypeptide of the invention and a further portion, wherein the said molecule is

not full-length *Xenopus* FAST1 or human FAST1 or a fragment thereof, mouse FAST2, *Xenopus* Milk, *Xenopus* Mixer or *Xenopus* Bix2. It is preferred that the said further portion confers a desirable feature on the said molecule; for example, the portion may be useful in detecting or isolating the molecule, or promoting cellular uptake of the molecule or the interacting polypeptide. The portion may be, for example, a biotin moiety, a radioactive moiety, a fluorescent moiety, for example a small fluorophore or a green fluorescent protein (GFP) fluorophore, as well known to those skilled in the art. The moiety may be an immunogenic tag, for example a Myc tag, as known to those skilled in the art or may be a lipophilic molecule or polypeptide domain that is capable of promoting cellular uptake of the molecule or the interacting polypeptide, as known to those skilled in the art, for example as characterised for a *Drosophila* polypeptide. Thus, the moiety may be derivable from the *Antennapedia* helix 3 (Derossi *et al* (1998) *Trends Cell Biol* 8, 84-87).

A particularly preferred molecule of the invention is Biotin.Aminohexanoicacid-RQIKIWFQNRRMKWKKLLMDFNNFPPNKTITPDMNVRIPPI, discussed in Example 1. The first 16 amino acids are from the helix 3 of *Antennapedia* which allows internalization of these peptides into live cells (Derossi *et al* 1998); the last 25 amino acids are codons 283-307 of Mixer.

A further aspect of the invention provides a nucleic acid (or polynucleotide) encoding or capable of expressing an interacting polypeptide or polypeptide containing PP(T/N)K of the invention. A still further aspect of the invention provides a nucleic acid complementary to a nucleic acid encoding or capable of expressing a polypeptide of the invention. Methods of preparing or isolating such a nucleic acid are well known to those skilled in the art.

The following methods of isolating a nucleic acid encoding an interacting polypeptide or polypeptide containing PP(T/N)K of the invention are given for purposes of illustration and are not considered to be exhaustive.

5

The polypeptide may be cleaved, for example using trypsin, cyanogen bromide, V8 protease formic acid, or another specific cleavage reagent. The digest may be chromatographed on a Vydac C18 column or subjected to SDS-PAGE to resolve the peptides. The N-terminal sequence of the peptides may then be  
10 determined using standard methods.

The sequences are used to isolate a nucleic acid encoding the peptide sequences using standard PCR-based strategies. Degenerate oligonucleotide mixtures, each comprising a mixture of all possible sequences encoding a part of the peptide  
15 sequences, are designed and used as PCR primers or probes for hybridisation analysis of PCR products after Southern blotting. mRNA prepared from cells in which the polypeptide may be expressed is used as the template for reverse transcriptase, to prepare cDNA, which is then used as the template for the PCR reactions.

20

Positive PCR fragments are subcloned and used to screen cDNA libraries to  
isolate a full length clone for the polypeptide.

Alternatively, the sequences of initial subcloned PCR fragments may be  
25 determined, and the sequence may then be extended by known PCR-based techniques to obtain a full length sequence.

Alternatively, the initial PCR sequence may be used to screen electronic databases of expressed sequence tags (ESTs) or other known sequences. By this means, related sequences may be identified which may be useful in isolating a full length sequence using the two approaches described above.

5

Sequences are determined using the Sanger dideoxy method. The encoded amino acid sequences may be deduced by routine methods.

Techniques used are essentially as described in Sambrook *et al* (1989) Molecular cloning, a laboratory manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Alternatively, antibodies may be raised against the polypeptide.

15 The antibodies are used to screen a  $\lambda$ gt11 expression library made from cDNA copied from mRNA from cells in which the polypeptide may be expressed.

Positive clones are identified and the insert sequenced by the Sanger method as mentioned above. The encoded amino acid sequence may be deduced by routine methods.

It will be appreciated that it may be desirable to express the polypeptide encoded by the isolated nucleic acid in order to determine that the polypeptide has the expected properties, for example that it is capable of interacting with a Smad polypeptide, for example Smad2 or Smad3.

25

The invention also includes a polynucleotide comprising a fragment of the



recombinant polynucleotide of the second aspect of the invention. Preferably, the polynucleotide comprises a fragment which is at least 10 nucleotides in length, more preferably at least 14 nucleotides in length and still more preferably at least 18 nucleotides in length. Such polynucleotides are useful as PCR primers.

The polynucleotide or recombinant polynucleotide may be DNA or RNA, preferably DNA. The polynucleotide may or may not contain introns in the coding sequence; preferably the polynucleotide is a cDNA.

A "variation" of the polynucleotide includes one which is (i) usable to produce a protein or a fragment thereof which is in turn usable to prepare antibodies which specifically bind to the protein encoded by the said polynucleotide or (ii) an antisense sequence corresponding to the gene or to a variation of type (i) as just defined. For example, different codons can be substituted which code for the same amino acid(s) as the original codons. Alternatively, the substitute codons may code for a different amino acid that will not affect the activity or immunogenicity of the protein or which may improve or otherwise modulate its activity or immunogenicity. For example, site-directed mutagenesis or other techniques can be employed to create single or multiple mutations, such as replacements, insertions, deletions, and transpositions, as described in Botstein and Shortle, "Strategies and Applications of *In Vitro* Mutagenesis," *Science*, **229**: 193-210 (1985), which is incorporated herein by reference. Since such modified polynucleotides can be obtained by the application of known techniques to the teachings contained herein, such modified polynucleotides are within the scope of the claimed invention.

Moreover, it will be recognised by those skilled in the art that the polynucleotide sequence (or fragments thereof) of the invention can be used to obtain other polynucleotide sequences that hybridise with it under conditions of high stringency. Such polynucleotides includes any genomic DNA. Accordingly, the polynucleotide of the invention includes polynucleotide that shows at least 55 per cent, preferably 60 per cent, and more preferably at least 70 per cent and most preferably at least 90 per cent homology with the polynucleotide identified in the method of the invention, provided that such homologous polynucleotide encodes a polypeptide which is usable in at least some of the methods described below or is otherwise useful.

Per cent homology can be determined by, for example, the GAP program of the University of Wisconsin Genetic Computer Group.

DNA-DNA, DNA-RNA and RNA-RNA hybridisation may be performed in aqueous solution containing between 0.1XSSC and 6XSSC and at temperatures of between 55°C and 70°C. It is well known in the art that the higher the temperature or the lower the SSC concentration the more stringent the hybridisation conditions. By "high stringency" we mean 2XSSC and 65°C. 1XSSC is 0.15M NaCl/0.015M sodium citrate. Polynucleotides which hybridise at high stringency are included within the scope of the claimed invention.

"Variations" of the polynucleotide also include polynucleotide in which relatively short stretches (for example 20 to 50 nucleotides) have a high degree of homology (at least 80% and preferably at least 90 or 95%) with equivalent stretches of the polynucleotide of the invention even though the overall homology between the two polynucleotides may be much less. This is because

important active or binding sites may be shared even when the general architecture of the protein is different.

A further aspect of the invention provides a replicable vector comprising a recombinant polynucleotide encoding an interacting polypeptide or a polypeptide containing PP(T/N)K of the invention. It will be appreciated that the said recombinant polynucleotide may encode an interacting polypeptide or polypeptide containing PP(T/N)K of the invention that is a fusion of an interacting polypeptide or polypeptide containing PP(T/N)K.

A variety of methods have been developed to operably link polynucleotides, especially DNA, to vectors for example *via* complementary cohesive termini. For instance, complementary homopolymer tracts can be added to the DNA segment to be inserted to the vector DNA. The vector and DNA segment are then joined by hydrogen bonding between the complementary homopolymeric tails to form recombinant DNA molecules.

Synthetic linkers containing one or more restriction sites provide an alternative method of joining the DNA segment to vectors. The DNA segment, generated by endonuclease restriction digestion as described earlier, is treated with bacteriophage T4 DNA polymerase or *E. coli* DNA polymerase I, enzymes that remove protruding, 3'-single-stranded termini with their 3'-5'-exonucleolytic activities, and fill in recessed 3'-ends with their polymerizing activities.

The combination of these activities therefore generates blunt-ended DNA segments. The blunt-ended segments are then incubated with a large molar excess of linker molecules in the presence of an enzyme that is able to catalyze

the ligation of blunt-ended DNA molecules, such as bacteriophage T4 DNA ligase. Thus, the products of the reaction are DNA segments carrying polymeric linker sequences at their ends. These DNA segments are then cleaved with the appropriate restriction enzyme and ligated to an expression vector that has been  
5 cleaved with an enzyme that produces termini compatible with those of the DNA segment.

Synthetic linkers containing a variety of restriction endonuclease sites are commercially available from a number of sources including International  
10 Biotechnologies Inc, New Haven, CN, USA.

A desirable way to modify the DNA encoding the polypeptide of the invention is to use the polymerase chain reaction as disclosed by Saiki *et al* (1988) *Science* 239, 487-491. This method may be used for introducing the DNA into a suitable  
15 vector, for example by engineering in suitable restriction sites, or it may be used to modify the DNA in other useful ways as is known in the art.

In this method the DNA to be enzymatically amplified is flanked by two specific primers which themselves become incorporated into the amplified DNA. The  
20 said specific primers may contain restriction endonuclease recognition sites which can be used for cloning into expression vectors using methods known in the art.

The DNA (or in the case of retroviral vectors, RNA) is then expressed in a  
25 suitable host to produce a polypeptide comprising the compound of the invention. Thus, the DNA encoding the polypeptide of the invention may be used in accordance with known techniques, appropriately modified in view of the

5 teachings contained herein, to construct an expression vector, which is then used to transform an appropriate host cell for the expression and production of the polypeptide of the invention. Such techniques include those disclosed in US Patent Nos. 4,440,859 issued 3 April 1984 to Rutter *et al*, 4,530,901 issued 23 July 1985 to Weissman, 4,582,800 issued 15 April 1986 to Crowl, 4,677,063 issued 30 June 1987 to Mark *et al*, 4,678,751 issued 7 July 1987 to Goeddel, 4,704,362 issued 3 November 1987 to Itakura *et al*, 4,710,463 issued 1 December 1987 to Murray, 4,757,006 issued 12 July 1988 to Toole, Jr. *et al*, 4,766,075 issued 23 August 1988 to Goeddel *et al* and 4,810,648 issued 7 March 10 1989 to Stalker, all of which are incorporated herein by reference.

15 The DNA (or in the case of retroviral vectors, RNA) encoding the polypeptide constituting the compound of the invention may be joined to a wide variety of other DNA sequences for introduction into an appropriate host. The companion DNA will depend upon the nature of the host, the manner of the introduction of the DNA into the host, and whether episomal maintenance or integration is desired.

20 Generally, the DNA is inserted into an expression vector, such as a plasmid, in proper orientation and correct reading frame for expression. If necessary, the DNA may be linked to the appropriate transcriptional and translational regulatory control nucleotide sequences recognised by the desired host, although such controls are generally available in the expression vector. The vector is then introduced into the host through standard techniques. Generally, not all of the 25 hosts will be transformed by the vector. Therefore, it will be necessary to select for transformed host cells. One selection technique involves incorporating into the expression vector a DNA sequence, with any necessary control elements,

that codes for a selectable trait in the transformed cell, such as antibiotic resistance. Alternatively, the gene for such selectable trait can be on another vector, which is used to co-transform the desired host cell.

- 5 Host cells that have been transformed by the recombinant DNA of the invention are then cultured for a sufficient time and under appropriate conditions known to those skilled in the art in view of the teachings disclosed herein to permit the expression of the polypeptide, which can then be recovered.
- 10 Many expression systems are known, including bacteria (for example *E. coli* and *Bacillus subtilis*), yeasts (for example *Saccharomyces cerevisiae*), filamentous fungi (for example *Aspergillus*), plant cells, animal cells and insect cells.

The vectors include a prokaryotic replicon, such as the ColE1 *ori*, for  
15 propagation in a prokaryote, even if the vector is to be used for expression in other, non-prokaryotic, cell types. The vectors can also include an appropriate promoter such as a prokaryotic promoter capable of directing the expression (transcription and translation) of the genes in a bacterial host cell, such as *E. coli*, transformed therewith.

- 20 A promoter is an expression control element formed by a DNA sequence that permits binding of RNA polymerase and transcription to occur. Promoter sequences compatible with exemplary bacterial hosts are typically provided in plasmid vectors containing convenient restriction sites for insertion of a DNA  
25 segment of the present invention.

Typical prokaryotic vector plasmids are pUC18, pUC19, pBR322 and pBR329

available from Biorad Laboratories, (Richmond, CA, USA) and pTrc99A and pKK223-3 available from Pharmacia, Piscataway, NJ, USA.

5 A typical mammalian cell vector plasmid is pSVL available from Pharmacia, Piscataway, NJ, USA. This vector uses the SV40 late promoter to drive expression of cloned genes, the highest level of expression being found in T antigen-producing cells, such as COS-1 cells.

10 An example of an inducible mammalian expression vector is pMSG, also available from Pharmacia. This vector uses the glucocorticoid-inducible promoter of the mouse mammary tumour virus long terminal repeat to drive expression of the cloned gene.

15 Useful yeast plasmid vectors are pRS403-406 and pRS413-416 and are generally available from Stratagene Cloning Systems, La Jolla, CA 92037, USA. Plasmids pRS403, pRS404, pRS405 and pRS406 are Yeast Integrating plasmids (YIps) and incorporate the yeast selectable markers *HIS3*, *TRP1*, *LEU2* and *URA3*. Plasmids pRS413-416 are Yeast Centromere plasmids (YCps).

20 The present invention also relates to a host cell transformed with a polynucleotide vector construct of the present invention. The host cell can be either prokaryotic or eukaryotic. Bacterial cells are preferred prokaryotic host cells and typically are a strain of *E. coli* such as, for example, the *E. coli* strains DH5 available from Bethesda Research Laboratories Inc., Bethesda, MD, USA, 25 and RR1 available from the American Type Culture Collection (ATCC) of Rockville, MD, USA (No ATCC 31343). Preferred eukaryotic host cells include yeast, insect and mammalian cells, preferably vertebrate cells such as

those from a mouse, rat, monkey or human fibroblastic and kidney cell lines. Yeast host cells include YPH499, YPH500 and YPH501 which are generally available from Stratagene Cloning Systems, La Jolla, CA 92037, USA. Preferred mammalian host cells include Chinese hamster ovary (CHO) cells  
5 available from the ATCC as CCL61, NIH Swiss mouse embryo cells NIH/3T3 available from the ATCC as CRL 1658, monkey kidney-derived COS-1 cells available from the ATCC as CRL 1650 and 293 cells which are human embryonic kidney cells. Preferred insect cells are Sf9 cells which can be transfected with baculovirus expression vectors.

10

Transformation of appropriate cell hosts with a DNA construct of the present invention is accomplished by well known methods that typically depend on the type of vector used. With regard to transformation of prokaryotic host cells, see, for example, Cohen *et al* (1972) *Proc. Natl. Acad. Sci. USA* **69**, 2110 and  
15 Sambrook *et al* (1989) *Molecular Cloning, A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY. Transformation of yeast cells is described in Sherman *et al* (1986) *Methods In Yeast Genetics, A Laboratory Manual*, Cold Spring Harbor, NY. The method of Beggs (1978) *Nature* **275**, 104-109 is also useful. With regard to vertebrate cells, reagents useful in  
20 transfecting such cells, for example calcium phosphate and DEAE-dextran or liposome formulations, are available from Stratagene Cloning Systems, or Life Technologies Inc., Gaithersburg, MD 20877, USA.

Electroporation is also useful for transforming and/or transfecting cells and is  
25 well known in the art for transforming yeast cell, bacterial cells, insect cells and vertebrate cells.



For example, many bacterial species may be transformed by the methods described in Luchansky *et al* (1988) *Mol. Microbiol.* 2, 637-646 incorporated herein by reference. The greatest number of transformants is consistently recovered following electroporation of the DNA-cell mixture suspended in 2.5X PEB using 6250V per cm at 25 $\mu$ FD.

Methods for transformation of yeast by electroporation are disclosed in Becker & Guarente (1990) *Methods Enzymol.* 194, 182.

10 Successfully transformed cells, ie cells that contain a DNA construct of the present invention, can be identified by well known techniques. For example, cells resulting from the introduction of an expression construct of the present invention can be grown to produce the polypeptide of the invention. Cells can be harvested and lysed and their DNA content examined for the presence of the  
15 DNA using a method such as that described by Southern (1975) *J. Mol. Biol.* 98, 503 or Berent *et al* (1985) *Biotech.* 3, 208. Alternatively, the presence of the protein in the supernatant can be detected using antibodies as described below.

In addition to directly assaying for the presence of recombinant DNA, successful  
20 transformation can be confirmed by well known immunological methods when the recombinant DNA is capable of directing the expression of the protein. For example, cells successfully transformed with an expression vector produce proteins displaying appropriate antigenicity. Samples of cells suspected of being transformed are harvested and assayed for the protein using suitable antibodies.

25 Thus, in addition to the transformed host cells themselves, the present invention also contemplates a culture of those cells, preferably a monoclonal (clonally

homogeneous) culture, or a culture derived from a monoclonal culture, in a nutrient medium.

A further aspect of the invention provides a method of making a polypeptide of the invention the method comprising culturing a host cell comprising a recombinant polynucleotide or a replicable vector which encodes said polypeptide, and isolating said polypeptide from said host cell. Methods of cultivating host cells and isolating recombinant proteins are well known in the art.

A further aspect of the invention provides an antibody capable of reacting with a polypeptide of the invention, in particular an antibody capable of reacting with an epitope comprising the amino acid sequence PP(T/N)K. Antibodies reactive towards the said polypeptide of the invention may be made by methods well known in the art. In particular, the antibodies may be polyclonal or monoclonal.

Suitable monoclonal antibodies may be prepared by known techniques, for example those disclosed in "Monoclonal Antibodies: A manual of techniques", H Zola (CRC Press, 1988) and in "Monoclonal Hybridoma Antibodies: Techniques and applications", J G R Hurrell (CRC Press, 1982), both of which are incorporated herein by reference. Other techniques for raising and purifying antibodies are well known in the art and any such techniques may be chosen to achieve the preparations useful in the methods claimed in this invention. Techniques for preparing antibodies are well known to those skilled in the art, for example as described in Harlow, ED & Lane, D "Antibodies: a laboratory manual" (1988) New York Cold Spring Harbor Laboratory.

Polyclonal antibodies may be prepared using methods well known in the art. In the case of both monoclonal and polyclonal antibodies, it is useful to use as immunogen any suitable polypeptide containing the PP(T/N)K motif. In particular with respect to the production of polyclonal antibodies it is useful to use polypeptides of between 10 and 30 amino acid residues containing the PP(T/N)K motif.

In a preferred embodiment of the invention, an antibody of the invention is capable of preventing or disrupting the interaction between a Smad polypeptide and a polypeptide comprising the amino acid sequence PP(T/N)K.

It will be appreciated that other antibody-like molecules may be useful in the practice of the invention including, for example, antibody fragments or derivatives which retain their antigen-binding sites, synthetic antibody-like molecules such as single-chain Fv fragments (ScFv) and domain antibodies (dAbs), and other molecules with antibody-like antigen binding motifs. Such antibody-like molecules are included by the term antibody as used below.

A further aspect of the invention provides a method of disrupting or preventing the interaction between a Smad polypeptide and a polypeptide (target polypeptide) that is (1) a transcription factor capable of interacting with the said Smad polypeptide and/or (2) a polypeptide capable of interacting with the said Smad polypeptide, the interaction requiring  $\alpha$ -helix2 of the said Smad polypeptide, the method comprising exposing the Smad polypeptide to an interacting polypeptide of the invention or an antibody of the invention. Alternatively, the Smad polypeptide may be exposed to a compound of the invention, as described below. It will be appreciated that the said polypeptide

capable of interacting with the said Smad polypeptide may interact with  $\alpha$ -helix 2 of the said Smad polypeptide; alternatively, the interaction may require  $\alpha$ -helix 2 but contact between the said polypeptide capable of interacting with the said Smad polypeptide and the said Smad polypeptide may occur at site in the said  
5 Smad polypeptide that is not part of  $\alpha$ -helix 2.

A further aspect of the invention provides a method of disrupting or preventing the interaction between a Smad polypeptide and a polypeptide (target polypeptide) which target polypeptide comprises the amino acid sequence  
10 PP(T/N)K the method comprising exposing the Smad polypeptide to an interacting polypeptide of the invention or an antibody of the invention. Alternatively, the Smad polypeptide may be exposed to a compound of the invention, as described below.

15 Preferences for the Smad polypeptide are as set out in relation to earlier aspects of the invention. It is particularly preferred that the Smad polypeptide is a naturally occurring Smad polypeptide, for example Smad2 or Smad3 or naturally occurring allelic variants thereof. It is still more preferred that the Smad polypeptide is a human Smad polypeptide, for example human Smad2 or human  
20 Smad3.

It is preferred that the antibody of the invention is capable of reacting with an epitope comprising the amino acid sequence PP(T/N)K.

25 The target polypeptide may be an interacting polypeptide of the invention, for example FAST3. It is preferred that the target polypeptide comprises the amino acid sequence PP(T/N)K. The target polypeptide may be FAST1, FAST2,

Mixer, Milk or Bix1, 2 or 3 or a fragment, variant, derivative or fusion thereof. It is preferred that the target polypeptide is a naturally occurring polypeptide or a fusion thereof.

- 5 The interaction between the Smad polypeptide and the target polypeptide and its disruption or prevention may be measured by any method of detecting/measuring a protein/protein interaction, as discussed further below and in Example 1. Suitable methods include yeast two-hybrid interactions, co-purification, ELISA, co-immunoprecipitation methods and bandshift assays.

10

The methods may be performed *in vitro*, either in intact cells or tissues, with broken cell or tissue preparations or at least partially purified components. Alternatively, they may be performed *in vivo*. The cells tissues or organisms in/on which the use or methods are performed may be transgenic. In particular  
15 they may be transgenic for the Smad interacting protein under consideration or for a further Smad interacting protein or Smad.

- A further aspect of the invention provides a method of identifying a polypeptide (interacting polypeptide) that is capable of interacting with a Smad polypeptide,  
20 for example Smad2 or Smad3, comprising examining the sequence of a polypeptide and determining that the polypeptide comprises the amino acid sequence PP(T/N)K. It is believed that the amino acid sequence PP(T/N)K is necessary and may be sufficient for interaction of a polypeptide with a Smad polypeptide, for example Smad2 or Smad3. Preferences for the Smad  
25 polypeptide are as given above. It may further be determined that an acidic amino acid residue is present at a position from 3 to 10, preferably 4 to 5 residues residues C-terminal of the amino acid sequence PP(T/N)K and/or a

proline residue is present at a position from 5 to 20 residues C-terminal of the amino acid sequence PP(T/N)K; these residues may also promote the interaction between the said interacting polypeptide and the Smad polypeptide. The acidic (negatively charged) amino acid residue is typically a glutamate or aspartate residue. The downstream proline and acidic, for example aspartate residues as described above may also be important for binding. It may further be determined that the residue immediately before (ie N-terminal of) the amino acid sequence PP(T/N)K is a hydrophobic residue, for example F, M or V. It may further be determined that the residue immediately after (ie C-terminal of) the amino acid sequence PP(T/N)K ie at position +1 is an S or T, which may be immediately followed by an I or V residue. It may further be determined that an acidic residue (for example glutamate or aspartate) present at position about +3 to about +10, preferably +4 or +5, is immediately followed by a hydrophobic residue, for example M, V or I.

Should the amino acid sequence of the said interacting polypeptide or the nucleotide sequence encoding the said interacting polypeptide not be known, they may be determined by methods well known to those skilled in the art, for example PCR-based cloning methods, as indicated above. It may be desirable to confirm that the interacting polypeptide identified by the method is capable of interacting with a Smad polypeptide, for example Smad2 or Smad3, using methods of detecting or measuring protein/protein interactions, as described above, for example using the interacting polypeptide expressed as described above.

The interacting polypeptide may also be useful in a screening assay for identifying a drug like compound that may inhibit the interaction between Smad2

or Smad3 and a polypeptide that interacts with Smad2 or Smad3 *in vivo*, for example a homologue of Milk, Mixer, other Mix family members, FAST1 and FAST2. It will be appreciated that the polypeptide may only interact with Smad2 or Smad3 when the Smad2 or Smad3 is in an activated state, for example following activation and/or phosphorylation as a consequence of TGF $\beta$  superfamily receptor activation, or wherein the N-terminal domain is not present or is truncated. It will be appreciated that the Smad2 or Smad3 may further interact with Smad4. It will be further appreciated that the Smad2 or Smad3 may interact or form a complex with more than one polypeptide that is not Smad4; for example, Smad2 or Smad3 may form a complex with Mixer, Milk and Smad4. Mixer and Milk may form a heterodimer.

A further aspect of the invention thus provides a method of identifying a compound capable of disrupting or preventing the interaction between a Smad polypeptide and a polypeptide (target polypeptide) that is (1) a transcription factor capable of interacting with the said Smad polypeptide and/or (2) a polypeptide capable of interacting with a Smad polypeptide, the interaction requiring  $\alpha$ -helix2 of the said Smad polypeptide and/or (3) a polypeptide comprising the amino acid sequence PP(T/N)K, the method comprising measuring the ability of the compound to disrupt or prevent the interaction between the Smad polypeptide and an interacting polypeptide of the invention.

The interaction between the Smad polypeptide and the interacting polypeptide and its disruption or prevention may be measured by any method of detecting/measuring a protein/protein interaction, as discussed in Example 1. Suitable methods include yeast two-hybrid interactions, co-purification, ELISA, co-immunoprecipitation methods and bandshift assays. Further suitable methods

may include Scintillation Proximity Assays, as well known to those skilled in the art. Examples of suitable methods may include bandshift assays looking for disruption of the endogenous FAST/Smad2/Smad4 ARF complex or disruption of the Mixer/GSTSmad2C interaction, as described in Example 1 and  
5 transcription assays in tissue culture cells in which expression of a reporter gene driven by a promoter with a binding site for (for example) Mixer is measured following treatment of the cells with TGF $\beta$ . Disruption or prevention of TGF $\beta$ -dependent transcription in the presence of the compound may be detected. The cells may be transiently transfected or may be a stable cell line capable of  
10 expressing Mixer (or other appropriate transcription factor) with an integrated reporter gene. The reporter gene may express luciferase or a green fluorescent protein (GFP), as well known to those skilled in the art. It will be appreciated that chip screening methods may be used. For example, arrays of cDNAs or oligonucleotides may be used in assessing expression of endogenous genes that  
15 are modulated by TGF $\beta$  and therefore for assessing effects of compounds on such expression.

The methods may be performed *in vitro*, either in intact cells or tissues, with broken cell or tissue preparations or at least partially purified components.  
20 Alternatively, they may be performed *in vivo*. The cells tissues or organisms in/on which the use or methods are performed may be transgenic. In particular they may be transgenic for the Smad interacting protein under consideration or for a further Smad interacting protein or Smad. Thus, a transgenic animal, for example a transgenic rodent, for example mouse or rat, amphibian, for example  
25 *Xenopus*, or insect, for example *Drosophila*, transgenic for the Smad interacting protein under consideration or for a further Smad interacting protein or Smad may be useful, for example in the screening methods of the invention.



It will be appreciated that screening assays which are capable of high throughput operation will be particularly preferred. Examples may include cell based assays, for example as described in Chen *et al* (1997) and protein-protein  
5 binding assays. An SPA-based (Scintillation Proximity Assay; Amersham International) system may be used. For example, beads comprising scintillant and a Smad polypeptide, for example Smad2 or a fragment (for example the MH2 domain) may be prepared. The beads may be mixed with a sample comprising the interacting polypeptide into which a radioactive label has been  
10 incorporated and with the test compound. Conveniently this is done in a 96-well format. The plate is then counted using a suitable scintillation counter, using known parameters for the particular radioactive label in an SPA assay. Only the radioactive label that is in proximity to the scintillant, ie only that bound to the interacting polypeptide that is bound to the Smad polypeptide anchored on the  
15 beads, is detected. Variants of such an assay, for example in which the Smad polypeptide is immobilised on the scintillant beads *via* binding to an antibody or antibody fragment, may also be used.

Other methods of detecting polypeptide/polypeptide interactions include  
20 ultrafiltration with ion spray mass spectroscopy/HPLC methods or other physical and analytical methods. Fluorescence Energy Resonance Transfer (FRET) methods, for example, well known to those skilled in the art, may be used, in which binding of two fluorescent labeled entities may be measured by measuring the interaction of the fluorescent labels when in close proximity to each other.

25

The compound may be a drug-like compound or lead compound for the development of a drug-like compound for each of the above methods of

identifying a compound. It will be appreciated that the said methods may be useful as screening assays in the development of pharmaceutical compounds or drugs, as well known to those skilled in the art.

5 The term “drug-like compound” is well known to those skilled in the art, and may include the meaning of a compound that has characteristics that may make it suitable for use in medicine, for example as the active ingredient in a medicament. Thus, for example, a drug-like compound may be a molecule that may be synthesised by the techniques of organic chemistry, less preferably by  
10 techniques of molecular biology or biochemistry, and is preferably a small molecule, which may be of less than 5000 daltons molecular weight. A drug-like compound may additionally exhibit features of selective interaction with a particular protein or proteins and be bioavailable and/or able to penetrate cellular membranes, but it will be appreciated that these features are not essential.

15

The term “lead compound” is similarly well known to those skilled in the art, and may include the meaning that the compound, whilst not itself suitable for use as a drug (for example because it is only weakly potent against its intended target, non-selective in its action, unstable, difficult to synthesise or has poor  
20 bioavailability) may provide a starting-point for the design of other compounds that may have more desirable characteristics.

It will be appreciated that the compound may be a polypeptide that is capable of competing with the interacting polypeptide of the invention for binding to the  
25 Smad polypeptide, and may be (1) a transcription factor capable of interacting with the said Smad polypeptide and/or (2) a polypeptide capable of interacting with a Smad polypeptide, the interaction requiring  $\alpha$ -helix2 of the said Smad

polypeptide and/or (3) a polypeptide comprising the amino acid sequence PP(T/N)K. Thus, it will be appreciated that a screening method as described above may be useful in identifying polypeptides that may interact with the Smad polypeptide.

5

Methods that may be useful in identifying polypeptides that may interact with the Smad polypeptide include yeast-2-hybrid, co-immunoprecipitation, ELISA, GST-pulldown, bandshift and transcription assays. Transcription assays may be performed *in vivo* or *in vitro*. For example, tissue culture cells may be used which comprise a reporter construct in which expression of the reporter gene is controlled a promoter comprising a binding site(s) for the putative interacting polypeptide. The effect of treating the cells with TGF $\beta$  on expression of the reporter gene may then be measured; TGF $\beta$ -dependent expression of the reporter gene may indicate that the putative interacting polypeptide is capable of being regulated by TGF $\beta$  and therefore may interact with the said Smad polypeptide. It will be appreciated that a transcription assay may be performed in a transgenic animal, for example a transgenic *Drosophila* or *Xenopus*.

10  
15

20

A further aspect of the invention is a kit of parts useful in carrying out a method, for example a screening method, of the invention. Such a kit may comprise a Smad polypeptide, for example Smad2 or Smad3 or a fragment either thereof and an interacting polypeptide, for example a polypeptide corresponding to amino acids 283 to 307 of Mixer.

25

A further aspect of the invention provides a compound identified by or identifiable by the screening method of the invention.

It will be appreciated that such a compound may be an inhibitor of the formation or stability of a complex of the Smad polypeptide used in the screen, for example Smad2 or Smad3, with interacting polypeptide(s), for example Smad4 and a transcription factor, for example FAST1, FAST2, Mixer, Milk or Bix2, and therefore ultimately of the activity of that complex, for example in promoting the transcription from a promoter to which the complex binds. The intention of the screen may be to identify compounds that act as modulators, for example inhibitors or promoters, preferably inhibitors of the activity of the complex, even if the screen makes use of a binding assay rather than an activity (for example transcriptional activity or DNA binding) assay. It will be appreciated that the inhibitory action of a compound found to bind the Smad or Smad interacting polypeptide may be confirmed by performing an assay of, for example, transcriptional or DNA binding activity in the presence of the compound.

A further aspect of the invention provides a compound identified by or identifiable by the screening method of the invention for use in medicine. A still further aspect of the invention provides an interacting polypeptide or polypeptide containing PP(T/N)K or molecule of the invention or nucleic acid of the invention or antibody of the invention for use in medicine.

The compound, interacting polypeptide, polypeptide containing PP(T/N)K, molecule, nucleic acid or antibody of the invention is suitably packaged and presented for use in medicine.

The aforementioned interacting polypeptide or molecule of the invention or nucleic acid of the invention or antibody of the invention or a formulation thereof may be administered by any conventional method including oral and

parenteral (e.g. subcutaneous or intramuscular) injection. The treatment may consist of a single dose or a plurality of doses over a period of time.

5 Whilst it is possible for an interacting polypeptide or molecule of the invention or nucleic acid of the invention or antibody of the invention to be administered alone, it is preferable to present it as a pharmaceutical formulation, together with one or more acceptable carriers. The carrier(s) must be "acceptable" in the sense of being compatible with the compound of the invention and not deleterious to the recipients thereof. Typically, the carriers will be water or  
10 saline which will be sterile and pyrogen free.

Thus, the invention also provides pharmaceutical compositions comprising the interacting polypeptide or molecule of the invention or nucleic acid of the invention or antibody of the invention and a pharmaceutically acceptable carrier.

15

As indicated above, the nucleic acid of the invention may be an antisense oligonucleotide, for example an antisense oligonucleotide directed against a nucleic acid encoding an interacting polypeptide of the invention, which may be a transcription factor comprising the amino acid sequence PP(N/T)K. It is  
20 preferred that the antisense oligonucleotide is directed against a nucleic acid encoding a human transcription factor.

---

Antisense oligonucleotides are single-stranded nucleic acid, which can specifically bind to a complementary nucleic acid sequence. By binding to the  
25 appropriate target sequence, an RNA-RNA, a DNA-DNA, or RNA-DNA duplex is formed. These nucleic acids are often termed "antisense" because they are complementary to the sense or coding strand of the gene. Recently, formation of

a triple helix has proven possible where the oligonucleotide is bound to a DNA duplex. It was found that oligonucleotides could recognise sequences in the major groove of the DNA double helix. A triple helix was formed thereby. This suggests that it is possible to synthesise a sequence-specific molecules  
5 which specifically bind double-stranded DNA *via* recognition of major groove hydrogen binding sites.

By binding to the target nucleic acid, the above oligonucleotides can inhibit the function of the target nucleic acid. This could, for example, be a result of  
10 blocking the transcription, processing, poly(A)addition, replication, translation, or promoting inhibitory mechanisms of the cells, such as promoting RNA degradations.

Antisense oligonucleotides are prepared in the laboratory and then introduced  
15 into cells, for example by microinjection or uptake from the cell culture medium into the cells, or they are expressed in cells after transfection with plasmids or retroviruses or other vectors carrying an antisense gene. Antisense oligonucleotides were first discovered to inhibit viral replication or expression in cell culture for Rous sarcoma virus, vesicular stomatitis virus, herpes simplex  
20 virus type 1, simian virus and influenza virus. Since then, inhibition of mRNA translation by antisense oligonucleotides has been studied extensively in cell-free systems including rabbit reticulocyte lysates and wheat germ extracts. Inhibition of viral function by antisense oligonucleotides has been demonstrated *in vitro* using oligonucleotides which were complementary to the AIDS HIV retrovirus  
25 RNA (Goodchild, J. 1988 "Inhibition of Human Immunodeficiency Virus Replication by Antisense Oligodeoxynucleotides", *Proc. Natl. Acad. Sci. (USA)* 85(15), 5507-11). The Goodchild study showed that oligonucleotides that were

most effective were complementary to the poly(A) signal; also effective were those targeted at the 5' end of the RNA, particularly the cap and 5' untranslated region, next to the primer binding site and at the primer binding site. The cap, 5' untranslated region, and poly(A) signal lie within the sequence repeated at the ends of retrovirus RNA (R region) and the oligonucleotides complementary to these may bind twice to the RNA.

Oligonucleotides are subject to being degraded or inactivated by cellular endogenous nucleases. To counter this problem, it is possible to use modified oligonucleotides, eg having altered internucleotide linkages, in which the naturally occurring phosphodiester linkages have been replaced with another linkage. For example, Agrawal *et al* (1988) *Proc. Natl. Acad. Sci. USA* **85**, 7079-7083 showed increased inhibition in tissue culture of HIV-1 using oligonucleotide phosphoramidates and phosphorothioates. Sarin *et al* (1988) *Proc. Natl. Acad. Sci. USA* **85**, 7448-7451 demonstrated increased inhibition of HIV-1 using oligonucleotide methylphosphonates. Agrawal *et al* (1989) *Proc. Natl. Acad. Sci. USA* **86**, 7790-7794 showed inhibition of HIV-1 replication in both early-infected and chronically infected cell cultures, using nucleotide sequence-specific oligonucleotide phosphorothioates. Leither *et al* (1990) *Proc. Natl. Acad. Sci. USA* **87**, 3430-3434 report inhibition in tissue culture of influenza virus replication by oligonucleotide phosphorothioates.

Oligonucleotides having artificial linkages have been shown to be resistant to degradation *in vivo*. For example, Shaw *et al* (1991) in *Nucleic Acids Res.* **19**, 747-750, report that otherwise unmodified oligonucleotides become more resistant to nucleases *in vivo* when they are blocked at the 3' end by certain capping

structures and that uncapped oligonucleotide phosphorothioates are not degraded *in vivo*.

A detailed description of the H-phosphonate approach to synthesising oligonucleoside phosphorothioates is provided in Agrawal and Tang (1990) *Tetrahedron Letters* **31**, 7541-7544, the teachings of which are hereby incorporated herein by reference. Syntheses of oligonucleoside methylphosphonates, phosphorodithioates, phosphoramidates, phosphate esters, bridged phosphoramidates and bridge phosphorothioates are known in the art.

See, for example, Agrawal and Goodchild (1987) *Tetrahedron Letters* **28**, 3539; Nielsen *et al* (1988) *Tetrahedron Letters* **29**, 2911; Jager *et al* (1988) *Biochemistry* **27**, 7237; Uznanski *et al* (1987) *Tetrahedron Letters* **28**, 3401; Bannwarth (1988) *Helv. Chim. Acta.* **71**, 1517; Crosstick and Vyle (1989) *Tetrahedron Letters* **30**, 4693; Agrawal *et al* (1990) *Proc. Natl. Acad. Sci. USA* **87**, 1401-1405, the teachings of which are incorporated herein by reference. Other methods for synthesis or production also are possible. In a preferred embodiment the oligonucleotide is a deoxyribonucleic acid (DNA), although ribonucleic acid (RNA) sequences may also be synthesised and applied.

The oligonucleotides useful in the invention preferably are designed to resist degradation by endogenous nucleolytic enzymes. *In vivo* degradation of oligonucleotides produces oligonucleotide breakdown products of reduced length. Such breakdown products are more likely to engage in non-specific hybridization and are less likely to be effective, relative to their full-length counterparts. Thus, it is desirable to use oligonucleotides that are resistant to degradation in the body and which are able to reach the targeted cells. The present oligonucleotides can be rendered more resistant to degradation *in vivo* by substituting one or more internal



artificial internucleotide linkages for the native phosphodiester linkages, for example, by replacing phosphate with sulphur in the linkage. Examples of linkages that may be used include phosphorothioates, methylphosphonates, sulphone, sulphate, ketyl, phosphorodithioates, various phosphoramidates, phosphate esters, bridged phosphorothioates and bridged phosphoramidates. Such examples are illustrative, rather than limiting, since other internucleotide linkages are known in the art. See, for example, Cohen, (1990) *Trends in Biotechnology*. The synthesis of oligonucleotides having one or more of these linkages substituted for the phosphodiester internucleotide linkages is well known in the art, including synthetic pathways for producing oligonucleotides having mixed internucleotide linkages.

Oligonucleotides can be made resistant to extension by endogenous enzymes by "capping" or incorporating similar groups on the 5' or 3' terminal nucleotides. A reagent for capping is commercially available as Amino-Link II<sup>TM</sup> from Applied BioSystems Inc, Foster City, CA. Methods for capping are described, for example, by Shaw *et al* (1991) *Nucleic Acids Res.* 19, 747-750 and Agrawal *et al* (1991) *Proc. Natl. Acad. Sci. USA* 88(17), 7595-7599, the teachings of which are hereby incorporated herein by reference.

20

~~A further method of making oligonucleotides resistant to nuclease attack is for~~  
~~them to be "self-stabilised" as described by Tang *et al* (1993) *Nucl. Acids Res.* 21,~~  
~~2729-2735 incorporated herein by reference. Self-stabilised oligonucleotides have~~  
~~hairpin loop structures at their 3' ends, and show increased resistance to~~  
~~degradation by snake venom phosphodiesterase, DNA polymerase I and fetal~~  
~~bovine serum. The self-stabilised region of the oligonucleotide does not interfere~~  
~~in hybridization with complementary nucleic acids, and pharmacokinetic and~~

stability studies in mice have shown increased *in vivo* persistence of self-stabilised oligonucleotides with respect to their linear counterparts.

It will be appreciated that antisense agents also include larger molecules which  
5 bind to said interacting polypeptide mRNA or genes and substantially prevent  
expression of said interacting polypeptide mRNA or genes and substantially  
prevent expression of said interacting polypeptide. Thus, expression of an  
antisense molecule which is substantially complementary to said interacting  
polypeptide is envisaged as part of the invention.

10 The said larger molecules may be expressed from any suitable genetic construct as  
is described below and delivered to the patient. Typically, the genetic construct  
which expresses the antisense molecule comprises at least a portion of the said  
interacting polypeptide coding sequence operatively linked to a promoter which  
15 can express the antisense molecule in the cell. Suitable promoters will be known  
to those skilled in the art, and may include promoters for ubiquitously expressed,  
for example housekeeping genes or for tissue-specific genes, depending upon  
where it is desired to express the antisense molecule.

20 Although the genetic construct can be DNA or RNA it is preferred if it is DNA.

Preferably, the genetic construct is adapted for delivery to a human cell.

Means and methods of introducing a genetic construct into a cell in an animal  
25 body are known in the art. For example, the constructs of the invention may be  
introduced into the cells by any convenient method, for example methods

involving retroviruses, so that the construct is inserted into the genome of the (dividing) cell.

Other methods involve simple delivery of the construct into the cell for expression therein either for a limited time or, following integration into the genome, for a longer time. An example of the latter approach includes liposomes (Nässander *et al* (1992) *Cancer Res.* **52**, 646-653). Other methods of delivery include adenoviruses carrying external DNA via an antibody-polylysine bridge (see Curiel *Prog. Med. Virol.* **40**, 1-18) and transferrin-polycation conjugates as carriers (Wagner *et al* (1990) *Proc. Natl. Acad. Sci. USA* **87**, 3410-3414). The DNA may also be delivered by adenovirus wherein it is present within the adenovirus particle. It will be appreciated that "naked DNA" and DNA complexed with cationic and neutral lipids may also be useful in introducing the DNA of the invention into cells of the patient to be treated.

Non-viral approaches to gene therapy are described in Ledley (1995) *Human Gene Therapy* **6**, 1129-1144. Alternative targeted delivery systems are also known such as the modified adenovirus system described in WO 94/10323 wherein, typically, the DNA is carried within the adenovirus, or adenovirus-like, particle. Michael *et al* (1995) *Gene Therapy* **2**, 660-668 describes modification of adenovirus to add a cell-selective moiety into a fibre protein. Mutant adenoviruses which replicate selectively in p53-deficient human tumour cells, such as those described in Bischoff *et al* (1996) *Science* **274**, 373-376 are also useful for delivering the genetic construct of the invention to a cell. Thus, it will be appreciated that a further aspect of the invention provides a virus or virus-like particle comprising a genetic construct of the invention. Other suitable viruses or virus-like particles include HSV, AAV, vaccinia and parvovirus.

A ribozyme capable of cleaving the interacting polypeptide RNA or DNA. A gene expressing said ribozyme may be administered in substantially the same and using substantially the same vehicles as for the antisense molecules. Ribozymes which may be encoded in the genomes of the viruses or virus-like particles  
5 herein disclosed are described in Cech and Herschlag "Site-specific cleavage of single stranded DNA" US 5,180,818; Altman *et al* "Cleavage of targeted RNA by RNase P" US 5,168,053, Cantin *et al* "Ribozyme cleavage of HIV-1 RNA" US 5,149,796; Cech *et al* "RNA ribozyme restriction endoribonucleases and methods", US 5,116,742; Been *et al* "RNA ribozyme polymerases,  
10 dephosphorylases, restriction endonucleases and methods", US 5,093,246; and Been *et al* "RNA ribozyme polymerases, dephosphorylases, restriction endoribonucleases and methods; cleaves single-stranded RNA at specific site by transesterification", US 4,987,071, all incorporated herein by reference.

15 The genetic constructs of the invention can be prepared using methods well known in the art.

A further aspect of the invention provides a method of modulating, for example enhancing or inhibiting, preferably inhibiting, activin or TGF $\beta$  signalling in a  
20 cell *in vitro* or *in vivo* wherein the cell is exposed to a polypeptide, molecule, nucleic acid, antibody or compound of the invention. It is preferred that the said polypeptide, molecule, nucleic acid, antibody or compound of the invention is able to enter the cell. Methods of optimising delivery to and uptake of such molecules by a cell are known to those skilled in the art and their use is  
25 envisaged here. The cell may be a tumour cell, for example a late stage tumour cell.

- A further aspect of the invention provides the use of a polypeptide, molecule, polynucleotide, compound or antibody of the invention in the manufacture of a medicament for treatment of a patient in need of modulation, preferably inhibition, of activin or TGF $\beta$  signalling. A further aspect of the invention
- 5 provides a method of treatment of a patient in need of modulation, preferably inhibition, of activin or TGF $\beta$  signalling wherein an effective amount of a polypeptide, polynucleotide, compound or antibody of the invention is administered to the patient.
- 10 A further aspect of the invention provides the use of a polypeptide, molecule, polynucleotide, compound or antibody of the invention in the manufacture of a medicament for treatment of a patient with cancer. A further aspect of the invention provides a method of treatment of a patient with cancer wherein an effective amount of a polypeptide, polynucleotide, compound or antibody of the
- 15 invention is administered to the patient.

For these and for following aspects of the invention it is preferred that the patient is mammalian. It is further preferred that the patient is human.

- 20 TGF $\beta$  is believed to be involved, for example, in scarring, tissue regeneration and kidney response to diabetes and therefore inhibition of TGF $\beta$  signalling ~~via the type-I and type-II receptors may be useful in medicine. Activin type-I and type-II receptors may be mediate~~ activins' roles in regulating endocrine cells from the reproductive system, promoters of erythroid differentiation and in
- 25 inducing axial mesoderm and anterior structures in vertebrates. Inhibins may have effects antagonistic to those of activins. BMP receptors may be involved in similar processes to TGF $\beta$  and activins, and particularly in bone growth and

maintenance. TGF $\beta$ s may be expressed in a wider range of tissues than other members of the superfamily, which may have more specialised roles.

5 TGF $\beta$  is also believed to be involved in carcinogenesis (see, for example  
Lawrence (1996), cited above) and therefore compounds that inhibit TGF $\beta$  and  
related receptor signalling may be useful in the treatment of cancer. Losses of  
Smad4 may be particularly associated with pancreatic and colon cancers; these  
cancers may not require TGF $\beta$  for progression. Breast cancer tumours are  
mentioned by Reiss (1997) *Oncol Res* 9, 447-457 as having high levels of TGF $\beta$   
10 associated with them and promoting tumour progression (see also Oft *et al*  
(1998) TGF $\beta$  signalling is necessary for carcinoma cell invasiveness and  
metastasis *Curr Biol* 8, 1243-1252).

A further aspect of the invention is the use of a polypeptide, molecule,  
15 polynucleotide, compound or antibody of the invention in the manufacture of a  
medicament for treatment of a patient in need of reducing extracellular matrix  
deposition, encouraging tissue repair and/or regeneration, tissue remodelling or  
healing of a wound, for example burn, injury or surgery, or reducing scar tissue  
formation arising from injury to the brain. A further aspect of the invention is a  
20 method of treatment of a patient in need of reducing extracellular matrix  
deposition, encouraging tissue repair and/or regeneration, tissue remodelling or  
healing of a wound, for example burn, injury or surgery, or reducing scar tissue  
formation arising from injury to the brain wherein an effective amount of a  
polypeptide, molecule, polynucleotide, compound or antibody of the invention is  
25 administered to the patient.

Extracellular matrix deposition is a term well known to those skilled in the art, and is described for example in Grande (1997) and Lawrence (1996), cited above. Extracellular matrix components include collagens, fibronectin, tenascin, glycosaminoglycans and proteoglycans. Deposition of such components may lead to rapid wound healing but may also lead to scarring, particularly in the brain. TGF $\beta$  may inhibit degradation of the extracellular matrix (for example by inhibiting production of proteases and stimulating the production of specific protease inhibitors).

- 10 It will be appreciated that the medicament may be applied before surgery. It will be appreciated that the injury may be mechanical injury. It is preferred that it is not reperfusion injury.

- A still further aspect of the invention is the use of a polypeptide, molecule, polynucleotide, compound or antibody of the invention in the manufacture of a medicament for treatment of a patient with or at risk of end-stage organ failure, pathologic extracellular matrix accumulation, a fibrotic condition, disease states associated with immunosuppression (such as different forms of malignancy, chronic degenerative diseases, and AIDS), diabetic nephropathy, tumour growth, kidney damage (for example obstructive neuropathy, IgA nephropathy or non-inflammatory renal disease) or renal fibrosis. A further aspect of the invention provides a method of treating a patient with or at risk of end-stage organ failure, pathologic extracellular matrix accumulation, a fibrotic condition, disease states associated with immunosuppression (such as different forms of malignancy, chronic degenerative diseases, and AIDS), diabetic nephropathy, tumour growth, kidney damage (for example obstructive neuropathy, IgA nephropathy or non-inflammatory renal disease) or renal fibrosis wherein an effective amount of a

polypeptide, molecule, polynucleotide, compound or antibody of the invention is administered to the patient.

5 The patient may alternatively have, or be at risk of, a form of a disorder of bone growth or homeostasis (such as osteoporosis), arthritis or atherosclerosis in which TGF $\beta$  or a related protein (for example an activin, inhibin or BMP) has been implicated, in causing or exacerbating the condition. The patient may be suffering from a TGF $\beta$ -related condition as reviewed in Roberts & Sport (1993) Physiological actions and clinical applications of transforming growth factor  $\beta$  10 (TGF $\beta$ ) *Growth Factors* 8, 1-9, for example hepatic cirrhosis, idiopathic pulmonary fibrosis, scleroderma, glomerulonephritis, certain forms of rheumatoid arthritis, schistosomiasis or proliferative vitreoretinopathy.

15 The polypeptide, molecule, polynucleotide, compound, antibody, composition or medicament of the invention may be administered in any suitable way, usually parenterally, for example intravenously, intraperitoneally or intravesically, in standard sterile, non-pyrogenic formulations of diluents and carriers. The polypeptide, molecule, polynucleotide, compound, antibody, composition or medicament of the invention of the invention may also be administered topically, 20 which may be of particular benefit for treatment of surface wounds. The polypeptide, molecule, polynucleotide, compound, antibody, composition or medicament of the invention may also be administered in a localised manner, for example by injection.

25 A further aspect of the invention provides a substantially pure complex comprising (1) a Smad2 or Smad3 polypeptide, (2) a Smad4 polypeptide and (3) a Mixer and/or Milk and/or Bix polypeptide. It will be appreciated that the



interactions between the components of the complex may be non-covalent interactions and that the complex may not be stable at non-physiological pH or salt concentrations. The complex may be stable and/or isolatable under conditions as described in Example 1 in which the complex may be detected by means of immunoprecipitation and/or band shift assays.

A further aspect of the invention provides a preparation comprising (1) Smad2 or Smad3 polypeptide, (2) a Smad4 polypeptide and (3) a Mixer and/or Milk and/or Bix polypeptide (in the form of a complex or otherwise) when combined with other components *ex vivo*, said other components not being all of the components found in the cell in which said (1) Smad2 or Smad3 polypeptide, (2) a Smad4 polypeptide and (3) a Mixer and/or Milk and/or Bix polypeptide (in the form of a complex or otherwise) are naturally found. The preparation may comprise a polypeptide that stabilises the preparation, for example bovine serum albumin or gelatin.

By "substantially pure" we mean that the complex is substantially free of other proteins. Thus, we include any composition that includes at least 30% of the protein content by weight as the said complex or its components, preferably at least 50%, more preferably at least 70%, still more preferably at least 90% and most preferably at least 95% of the protein content is the said complex or its components.

Thus the substantially pure complex may include a contaminant wherein the contaminant comprises less than 70% of the composition by weight, preferably less than 50% of the composition, more preferably less than 30% of the composition, still more preferably less than 10% of the composition and most preferably less than 5% of the composition by weight.

The substantially pure said complex may be combined with other components *ex vivo*, said other components not being all of the components found in the cell in which said complex is naturally found.

5

The invention will now be described in more detail by reference to the following Figures and Examples.

### Figure legends

10 Figure 1. Activin-responsive transcription via the goosecoid DE.

(A) Activin-responsive transcription via the DE is partially dependent on new protein synthesis. One-cell embryos were injected with REF-globin internal control together with globin reporters driven by the minimal  $\gamma$ -actin promoter ( $\gamma$ A), or by multiple copies of the DE or ARE upstream of the minimal promoter.

15 Animal caps, cut at St 8 were cultured for 6 h  $\pm$  activin in the absence or presence of 5  $\mu$ g/ml cycloheximide. Globin transcripts from reporter genes (Test-globin) or the internal control (REF-globin) were detected by RNase protection (Howell and Hill, 1997). Transcriptional activation was calculated as a ratio of the levels of Test-globin to REF-globin. Activin-induced transcription is expressed as fold  
20 inductions. In close agreement with the data shown in this experiment, a similar independent experiment measuring the activin inducibility of the DE gave a 22.1-fold induction in the absence of cycloheximide and 6.1-fold in the presence of cycloheximide.

(B) An activin-inducible factor (DEBP) binds the goosecoid DE. Whole cell  
25 extracts prepared from St 8 or St 11 embryos or St 11 embryos overexpressing activin, were analysed by bandshift assay using the single DE probe. Activin-

inducible factor, DEBP is indicated. Competitor oligonucleotides were used at a 50-fold molar excess over probe where indicated. Below, sequences of wild-type DE and mutant oligonucleotides, where only the altered nucleotides are indicated. The paired-like homeodomain binding site comprising 2 inverted TAAT motifs (Wilson *et al* 1993) is denoted by arrows; a third homeodomain binding site at the 3' end is also indicated by an arrow. Thick dotted line, sequence reminiscent of a half-site for the T-box protein, brachyury (AGGTGTGAAATT) (Kispert *et al* 1995), and overlapping this (underlined) is an almost perfect binding site for the ZFH-1 family of zinc finger homeodomain proteins (AGGTGAGCAA) (Funahashi *et al* 1993).

(C) Formation of DEBP requires new protein synthesis. Extracts were made from uninjected St 8 embryos (lane 1), St 10.5 embryos (lanes 2, 3), or St 10.5 embryos overexpressing activin (lanes 4,5) and analysed by bandshift using the DE probe. Where indicated, embryos had been pre-incubated in cycloheximide before St 8.

Figure 2. The effector domain of Smad2 interacts with DEBP.

Whole cell extracts prepared from either St 8 embryos (lanes 1-4), St 10.5 embryos (lanes 5-8) or St 10.5 embryos overexpressing activin (lanes 9-12) were analysed by bandshift using the DE probe. Extracts were mixed with either purified GST protein (100 ng) (lanes 4, 8, 12) or 2 concentrations (20 ng and 100 ng) of purified GSTSmad2C (lanes 2,3, 6,7, 10,11) prior to addition of probe. Open arrow, DEBP; black arrow, GSTSmad2C associated with DEBP.

Figure 3. Homeodomain proteins, Mixer and Milk, but not Mix.1, interact with Smad2C

(A) Overexpression of Mixer and Milk in *Xenopus* embryos mimics the activin induction of DEBP. Whole cell extracts were prepared from St 10.5 embryos or embryos injected at the 1-cell stage with mRNA encoding myc-tagged Mixer, Milk, Mix.1, or activin, and DE-binding activity was assayed by bandshift. Anti-myc antibody or purified GSTSmad2C were added where indicated. Open arrow, DEBP; gray arrow, supershifted complexes.

(B) Interaction of GSTSmad2C with members of the Mix family and Fast-1. In-vitro translated Mixer, Milk, Mix.1 and Fast.1 were assayed by bandshift for their interaction with purified GSTSmad2C or GST using the appropriate radiolabelled DE or ARE probes. Open arrow, transcription factors complexed with probe; black arrow, ternary complex with GSTSmad2C.

Figure 4. Characterization of the Smad Interaction Motif (SIM)

(A) Schematics of Mix.1, Mixer and Milk, with the conserved homeodomains and a C-terminal acidic domain indicated. Black box; a region conserved in Milk and Mixer, also present in *Xenopus* Fast-1 and mouse Fast-2 (expanded below where the black line denotes the boundaries of the conserved sequences). Black shading, identical amino acids; gray shading, similar amino acids. The numbers indicate the positions of these amino acids in the full length sequences of the individual proteins.

(B) C-terminal deletion mutants of Mixer, Milk and Fast.1 (schematized below) were produced *in vitro* and their interaction with GSTSmad2C assayed by bandshift using the DE or ARE probe as appropriate. Complexes of transcription factors and probe are indicated; black arrow, ternary complex with GSTSmad2C. SIM, Smad interaction motif and DNA-binding domains are indicated. Note that

Milk gives rise to two complexes, both of which shift with GSTSmad2C, which correspond to a dimer of Milk and a higher order complex.

(C) Mutation of the prolines in the PP(T/N)K core motif abolishes the interaction with Smad2C. Full length Mixer or a mutant derivative (Mixer PP mut), in which the 2 prolines in the PP(T/N)K containing motif are mutated to alanines, were produced *in vitro* and assayed for interaction with GSTSmad2C by bandshift using the DE probe.

(D) Interaction of Mixer and Milk and Fast-1 with Smad2C in solution. [<sup>35</sup>S]-labelled transcription factors as indicated were incubated with Sepharose-bound GST (lanes 2,5,8,11,14) or GSTSmad2C (lanes 3,6,9,12,15) and bound protein was visualized by SDS-PAGE and autoradiography. A fraction of input protein was analysed for comparison (lanes 1,4,7,10,13).

Figure 5. The Smad interaction motif is sufficient to interact with Smad2.

(A) A peptide containing the SIM of Mixer competes specifically for interaction of Mixer with Smad2C. *In vitro*-translated Mixer was incubated with DE probe alone (lane 1) or in the presence of 1 or 10 pmoles of wild type peptide (lanes 2,3) or mutant peptide (lanes 4,5). GSTSmad2C (20 ng) was included in the reactions in lanes 6-14, with the addition of 0.3, 1, 3 or 10 pmoles of wild type peptide (lanes 7-10) or mutant peptide (lanes 11-14). Mixer complexed with probe is indicated; black arrow, ternary complex with GSTSmad2C.

(B) A peptide containing the SIM of Mixer specifically disrupts the formation of ARF.

Whole cell extracts made from activin-injected St 8 embryos were analysed by bandshift assay with the ARE probe in the absence (lane 1) or presence of 10, 30,

60, 100, 200 pmoles wild type peptide (lanes 2-6) or mutant peptide (lanes 7-11).

The endogenous ARF complex is indicated.

For peptide sequences see Experimental Procedures.

5 Figure 6. Mixer and Milk interact with activated Smads *in vivo*

(A) Mixer forms a ligand-dependent complex with Smad2 and Smad4 in solution.

10 Extracts were prepared from NIH3T3 cells transfected with myc-Smad2, myc-Smad4 and either Flag-Fast-1, Flag-Mixer, or a Flag-tagged mutant derivative (Mixer PP mut), which had been incubated  $\pm$  TGF- $\beta$ 1 (2 ng/ml) for 1h. Extracts were assayed either by immunoprecipitation of complexes with anti-Flag antibody followed by Western blotting with anti-Myc antibody (top panel), or Western blotting the whole extract with anti-Flag antibody (middle panel) or with anti-Myc antibody (bottom panel).

15 (B and C) Fast-1 and Mixer form ligand-dependent complexes on DNA with endogenous Smad2 and Smad4. Extracts were prepared from NIH3T3 cells transfected with Flag-tagged Fast-1, Mixer or Mixer (PP mut), which had been incubated  $\pm$  TGF- $\beta$ 1 (2 ng/ml) for 1h. Extracts were analysed by bandshift assay on the ARE (B) or DE (C) probe. Anti-flag, anti-Smad2 or anti-Smad4 antibodies were included in the binding reactions where indicated. In (B) ARF and antibody-supershifted ARF are indicated. In (C), Mixer or Mixer (PP mut) bound to probe, 20 the Mixer-Smad complex and antibody-supershifted Mixer-Smad complex are indicated.

Figure 7. Mixer and Milk mediate TGF- $\beta$ -dependent transcriptional activation  
25 via the DE

(A). NIH3T3 cells were transfected with the CAT reporters, and plasmids expressing transcription factors, Smad2 and Smad4 as indicated. Cells were cultured  $\pm$  TGF- $\beta$ 1 (2 ng/ml) for 8 hr. Cells were harvested and CAT activity measured relative to lacZ activity from the internal control. The data are from a representative experiment, and similar results were obtained in at least three further independent experiments.

(B) Mixer mediates TGF- $\beta$  dependent transcriptional activation via the DE in the absence of protein synthesis. NIH3T3 cells were transfected with the (DE)<sub>4</sub>-globin reporter and REF-globin internal control with or without Mixer expression plasmid. Cells were cultured  $\pm$ TGF- $\beta$ 1 (2 ng/ml) for 4 hr in the absence or presence of 50  $\mu$ g/ml cycloheximide. Globin transcripts from the reporter genes (test-globin) or the internal control (REF-globin) were detected by RNase protection and quantitated as in Figure 1.

Figure 8. The temporal and spatial expression patterns of *Mixer* and *Milk* in *Xenopus* embryos makes them good candidates for mediating transcription of *goosecoid* in response to an endogenous activin-like signal.

(A) Co-expression of *goosecoid* with *Mixer* and *Milk* at early gastrula stages. *Xenopus* embryos were fixed at St 10.25 and processed for in situ hybridization with probes against *goosecoid* (*Gsc*), *Mixer* or *Milk* either singly (left panels) or sequentially (right panels). Arrowhead, dorsal lip. *Gsc* mRNA is visualized with deep purple stain. *Mixer* and *Milk* mRNA are visualized with a turquoise stain. In the double in situs the overlapping turquoise *Mixer* or *Milk* staining with the purple *Gsc* staining is evident as dark blue staining in dorsal marginal zone (above the dorsal lip). The weak purple background of these embryos is non-specific staining.

(B) Temporal expression patterns of *Mixer*, *Milk* and *goosecoid* in *Xenopus* embryos.

Time course of expression of *goosecoid* (*Gsc*), *Mixer*, *Milk* and the *FGF receptor* (*FGFR*) assayed by RNase protection. Embryos were sampled at St 8 and subsequent times indicated. In lanes 9-16 the embryos had been pre-incubated with cycloheximide from 30 min before St 8. The *Milk* probe also detects a highly related mRNA, *Milk-related* which is likely to be *Bix3*, which also has a very well conserved PP(T/N)K-containing SIM (see text; Tada *et al* 1998).

(C) A model showing that TGF- $\beta$ /activ in activated Smads translocate to the nucleus, where they interact with homeodomain transcription factors, Mixer and Milk through the SIM to activate transcription.

(D) A model describing the proposed role of the Mixer/Milk-Smad complexes in the formation of mesoderm and endoderm in early *Xenopus* embryos. The black arrows denote induction of gene expression; the gray arrows denote activation of protein complexes. Milk-related protein/Milk-Smad complexes are involved in the initiation of transcription of meso-endodermal genes and Mixer/Milk/Bix-Smad complexes are involved in the maintenance of gene expression. For discussion, see text.

Figure 9. Mixer and Milk mediate TGF- $\beta$ -dependent transcriptional activation via a single DE.

NIH3T3 cells were transfected with a CAT reporter gene driven by a single copy of the *goosecoid* DE, and plasmids expressing transcription factors Mixer, Mixer (PP mut), Milk and Fast-1 as shown. Cells were cultured  $\pm$  TGF- $\beta$ 1 (2 ng/ml) for 8 hr. Cells were harvested and CAT activity measured. The data are from a



representative experiment, and similar results were obtained in two further independent experiments.

Figure 10. An activin-inducible factor (DEBP) binds the paired-like homeodomain binding site of the *goosecoid* DE.

Whole cell extracts prepared from St 8 or St 11 embryos or St 11 embryos overexpressing activin, were analysed by bandshift assay either on the wild type DE probe, or mutant DE probes as indicated. Activin-inducible factor, DEBP is indicated. Below, sequences of wild-type DE and mutant oligonucleotides, where only the altered nucleotides are indicated. The paired-like homeodomain binding site comprising 2 inverted TAAT motifs (Wilson *et al* 1993) is denoted by arrows; a third homeodomain binding site at the 3' end is also indicated by an arrow. Thick dotted line, sequence reminiscent of a half-site for the T-box protein, brachyury (AGGTGTGAAATT) (Kispert *et al* 1995), and overlapping this (underlined) is an almost perfect binding site for the ZFH-1 family of zinc finger homeodomain proteins (AGGTGAGCAA) (Funahashi *et al* 1993).

Figure 11. Mapping the Mixer interaction domain in Smad2.

(Left panel) In-vitro translated Mixer was tested in a bandshift assay, using radiolabelled DE as probe, for its ability to interact with different Smad2 effector domain mutants, produced bacterially as GST fusion proteins. Mixer complexed with probe is indicated; black arrow, ternary complex with GSTSmad2C derivatives. (Right panel) Helix 2 of the Smad2 effector domain is required for the interaction with Mixer. The assay was as above. The effector domain of Smad1 does not interact with Mixer. In the mutant GSTSmad2C (H2 swap) Helix 2 from Smad1 replaces Helix 2 of Smad2. This mutant contains only 4 amino acid

changes relative to GSTSmad2C (Shi *et al* 1997) and no longer interacts with Mixer.

Figure 12. Amino acid sequence of human and *Xenopus* Smad2 and Smad3.

5 *Xenopus* Smad3 is a novel Smad polypeptide.

Figure 13. Amino acid and nucleotide sequences of *Xenopus* Bix1, Bix2, Bix3, Bix4, mixer, Milk, Mix.1, Mix.2, FAST1 and FAST3; human FAST1; mouse FAST2. References for these sequences are given in the text. FAST3 is a novel

10 FAST polypeptide.

Figure 14. Amino acid sequence of Chicken Mix and alignment with *Xenopus* Mix.1 amino acid sequence. References for the sequences are given in the text.

## 15 **Example 1: Homeodomain Transcriptional Partners For Smads**

Smads transduce TGF- $\beta$  signals and participate in transcriptional regulation. We now identify paired-like homeodomain transcription factors of the *Xenopus* Mix family as new partners for activated Smads. We identify a DE-binding protein (DEBP) in *Xenopus* embryos which is synthesized in response to activin and its binding to the paired-like homeodomain site in the DE correlates with activin-induced transcription. DEBP specifically interacts with the effector domain of the activin-activated Smad, Smad2. We demonstrate that two members of the *Xenopus* Mix family of paired-like homeodomain transcription factors, Mixer (Henry and Melton, 1998) and Milk (Ecochard *et al* 1998) precisely mimic the activity of endogenous DEBP. We demonstrate that Mixer and Milk, but not a third family member, Mix. 1 (Rosa, 1989), directly interact with activin/TGF- $\beta$ -activated

Smad2. This allows recruitment of Smad4 to form an activin/TGF $\beta$ -inducible complex that mediates transcriptional activation via the goosecoid DE, the activin-responsive element of the *Xenopus* goosecoid promoter. We have identified a short motif in the C-terminal region of Mixer and Milk, characterized by the sequence PP(T/N)K, which is necessary and sufficient for interaction with the MH2 domain of Smad2. This Smad interaction motif (SIM) is also conserved in the C-terminal regions of the unrelated Smad2-interacting forkhead transcription factors, Fast-1 and Fast-2. Furthermore, we show that Mixer and Milk are expressed in the same cells of the *Xenopus* embryo that express goosecoid, strongly suggesting they are responsible for regulating transcription of goosecoid *in vivo* in response to the endogenous activin-like signals through their interactions with Smads. Our data lead us to propose a model for meso-endoderm formation in *Xenopus* in which these homeodomain transcription factor/Smad complexes play a central role in initiating and maintaining transcription in response to endogenous TGF $\beta$ /activin-like signals.

## Results

### Activin induced transcription via the distal element of the goosecoid promoter is partly dependent on new protein synthesis

The distal element (DE) in the *Xenopus* *goosecoid* promoter is a cis-acting element necessary and sufficient to activate transcription in response to activin (Watabe *et al* 1995). We first investigated activin-stimulated transcription via the DE in animal cap assays (Howell and Hill, 1997), and compared it with the transcriptional response of the ARE from the *Xenopus* *Mix.2* promoter, which has a completely different sequence and is known to be controlled by the Fast-

1/Smad2/Smad4 complex, ARF (Huang *et al* 1995; Chen *et al* 1996; Chen *et al* 1997). We used globin reporter genes with four copies of the DE or three copies of the ARE linked to a minimal promoter and measured transcription by RNase protection assay, quantitating it relative to the activity of a co-injected

5 constitutively active reference globin gene (Howell and Hill, 1997). To get an accurate value for transcription from the TEST-globin plasmid, the amount of transcript from TEST-Globin has to be divided by the amount of transcript from the REF-globin. REF-globin acts as an internal control for injection efficiency, RNA extraction efficiency and as a loading control. The minimal promoter was

10 unresponsive to activin (Figure 1A, left panel). The reporter driven by four DEs responded to activin strongly, and some of this induction was lost in the presence of the protein synthesis inhibitor, cycloheximide (middle panel). The ARE in contrast gave a much higher basal level of transcription, and the activin induction was weaker. As expected, this induction was completely insensitive to

15 cycloheximide (right panel), consistent with it being mediated by the maternal transcription factor complex, ARF.

From this experiment we conclude that the transcriptional response of the DE to activin has two components: a direct induction mediated by maternal factors,

20 which is insensitive to cycloheximide and a maintenance phase which requires new protein synthesis. A similar behaviour was recently proposed for the related activin-responsive sequence in the zebrafish *gooseoid* promoter (McKendry *et al* 1998).

25 **The *gooseoid* DE binds an activin-inducible factor, DEBP**

We used bandshift assays with a radiolabelled single DE oligonucleotide as probe to identify DE-binding factors in the embryo that might be responsible for activin-induced transcription. The DE-binding factor which displayed the expected behaviour is DEBP (DE-binding protein; Figure 1B, lanes 1-3, open arrow). It was  
 5 absent in extracts prepared from Stage 8 embryos, which are transcriptionally inactive (lane 1). It was present at low levels in extracts from Stage 11 embryos in which endogenous activin-like signaling pathways are operating (lane 2; (Sun *et al* 1999), and highly induced in Stage 11 embryos overexpressing activin (lane 3). Binding of this complex to the DE was specific since it was competed by excess  
 10 homologous unlabelled probe (lanes 4-6). This complex is probably the same as GAEBP1, shown to bind the related activin-responsive sequence in the zebrafish *goosecoid* promoter (McKendry *et al* 1998).

The DE contains binding sites for several different DNA-binding proteins: a  
 15 consensus for a paired-like homeodomain protein at its 5' end, consisting of two inverted TAAT motifs separated by 3 nucleotides (Wilson *et al* 1993; McKendry *et al* 1998; arrows, Figure 1B); an additional homeodomain core binding site at the 3' end (arrow); a sequence reminiscent of a half site for the T-box protein, brachyury (dotted line; Kispert *et al* 1995), and overlapping this, a binding site for  
 20 the ZFH-1 family of zinc finger homeodomain proteins (underlined; Funahashi *et al* 1993). We performed competitions with various DE mutants to determine which of these binding sites was required for DEBP binding.

DE m1, which is mutated in the paired-like homeodomain binding site (Watabe *et al* 1995), competed very poorly for binding (lanes 7-9), indicating that this site is  
 25 required. This mutant was also completely inactive in activin-responsive transcription assays (data not shown; Watabe *et al* 1995; McKendry *et al* 1998),

indicating that this paired-like homeodomain binding site is absolutely required for activin-responsive transcription. DE m2, which is mutated in all three homeodomain binding sites, did not compete for DEBP binding at all (lanes 10-12). DE m3, in contrast, which is mutated in the T-box and ZFH-1 and the 3' homeodomain binding site, competed efficiently for binding (lanes 13-15), indicating that these sites were not required. The ARE did not compete for DEBP binding (lanes 16-18). Bandshift assays using these mutants as probes were consistent with these conclusions (data not shown).

10 Since the activin-responsive transcription via the DE is partly dependent on new protein synthesis, we asked whether the activin-inducible DEBP also required new protein synthesis for its formation. Indeed, preincubation of the embryos with cycloheximide prior to initiation of zygotic transcription abolished formation of DEBP either in Stage 10.5 embryos or Stage 10.5 embryos overexpressing activin  
15 as assayed by bandshift (Figure 1C).

Thus the activin-inducible DEBP binds to the paired-like homeodomain binding site of the DE. The fact that the integrity of this binding site is absolutely required for all the activin-responsive transcription of the DE strongly suggests that DEBP is involved in this (McKendry *et al* 1998). The observation that activin induction of DEBP requires new protein synthesis, indicates that DEBP is most likely to mediate the maintenance phase of transcription of the DE in response to activin. However low levels of maternal DEBP might mediate the component of activin-responsive transcription of the DE that does not require new protein synthesis  
25 (McKendry *et al* 1998; see Discussion).

### The activin-inducible DEBP can interact with Smad2

Activin signals are transduced from activated receptors to the nucleus via a complex of activated Smad2 and Smad4 (Massagué, 1998). We therefore asked whether DEBP might correspond to a Smad/transcription factor complex, by analogy with the Fast-1/Smad complex, ARF. An antibody specific for Smad2 (Nakao *et al* 1997) did not supershift DEBP, indicating that DEBP did not contain endogenous Smad2 and was thus unlikely to be a Smad/transcription factor complex (data not shown). The same Smad2 antibody however could efficiently supershift the ARF complex (data not shown; see Figure 6B).

An alternative possibility was that DEBP was a DNA-binding protein that did interact with activated Smads, but the resulting Smad/DEBP complex was not detectable in our bandshift assays. We therefore investigated whether DEBP could interact with the effector MH2 domain of Smad2 (Smad2C), which is the domain of Smad2 that interacts with Fast-1 in the ARF complex (Chen *et al* 1997; Liu *et al* 1997). Indeed, purified Smad2C, bacterially-expressed as a GST-fusion protein (GSTSmad2C) stoichiometrically supershifted DEBP generated by the endogenous activin-like signals (Figure 2, lanes 5-7) or that generated in response to high levels of activin signaling (lanes 9-11). This was specific, as GST alone had no effect (lanes 8, 12). As expected GSTSmad2C alone did not bind the DE probe, as seen by the lack of binding activity when added to Stage 8 embryo extracts which do not contain DEBP (lanes 1-3). Thus the supershifts (lanes 6,7,10,11) arise from binding of GSTSmad2C to DEBP, and we conclude that DEBP can interact with the effector domain of Smad2.

### Identification of Smad2-interacting transcription factors that mimic DEBP

The activin-inducible DEBP therefore appears to act as a platform for recruiting Smad2. UV-cross-linking experiments indicated that DEBP corresponded to a monomer of approximately 45-50 kDa (data not shown). In addition, DEBP binds the paired-like homeodomain binding site of the DE and is synthesized in response to activin. A group of transcription factors with precisely these properties are the paired-like homeodomain proteins of the Mix family. There are seven family members: Mix.1 and the highly related Mix.2 (Rosa, 1989; Vize, 1996), Mixer (Henry and Melton, 1998), Milk (Ecochard *et al* 1998), also called Bix2 (Tada *et al* 1998) and three other Bix genes which are highly related to Milk (Tada *et al* 1998). They all have molecular weights of approximately 44 kDa, are first expressed at the mid to late blastula stage of *Xenopus* embryogenesis and their expression is known to be induced by activin signaling.

We asked whether overexpression in *Xenopus* embryos of three different Mix family members, Mix.1, Mixer and Milk, could mimic the activity of DEBP, both in DNA-binding specificity and in their ability to interact with Smad2C. Overexpression of myc-tagged Mixer, Milk or Mix.1 alone gave rise to protein/DNA complexes that co-migrated with the activin-induced DEBP (Figure 3A, compare lanes 1, 4, 7, 10 with 13). These protein/DNA complexes could be supershifted with the anti-myc antibody (lanes 5,8,11) indicating the myc-tagged proteins are constituents. Strikingly, only Mixer and Milk have the ability to interact with GSTSmad2C, as shown for endogenous DEBP (compare lanes 6,9, with 3,15). Mix.1 could not associate with GSTSmad2C (lane 12).

We performed an analogous interaction experiment using transcription factors produced *in vitro* by coupled transcription/translation with identical results (Figure



3B, lanes 1-9). As a control for the supershift bandshift assay we also tested the known Smad2-interacting protein, Fast-1 (Chen *et al* 1996), which can be supershifted by GSTSmad2C, but not by GST alone (Figure 3B, lanes 10-12).

- 5 Thus Mixer and Milk, but not Mix.1, interact with the effector domain of Smad2, and are therefore good candidates for endogenous DEBP.

#### **Sequences of Smad2 required for interaction with DEBP, Mixer, Milk and Fast-1**

- 10 We next investigated the sequences in Smad2 required to interact with Mixer, Milk, Fast-1 and endogenous DEBP by assaying a series of Smad2C deletion mutants in the supershift bandshift assay described above (Table 1). Deletion of the phosphorylation sites in the SSMS motif at the extreme C-terminus of Smad2 had no effect on binding to any of the transcription factors (mutant 198-463).
- 15 Analysis of further N- and C-terminal deletions indicated that the integrity of most of the Smad2 MH2 domain was required for binding to the transcription factors (Table 1). Interestingly Mixer behaved identically to the endogenous DEBP in its interaction with Smad2, whilst Milk behaved like Fast-1 and required additional residues at the C-terminal domain of Smad2C (Table 1, compare mutants 198-445,
- 20 198-440, and 198-426). The interaction of the transcription factors with Smad2C was specific, since the equivalent C-terminal region of the BMP-activated Smad, Smad1 (GSTSmad1C) could not interact.

- The region of Smad2 thought to contact Fast-1 has previously been elucidated and
- 25 is the  $\alpha$ -helix-2 (Chen *et al* 1998). We therefore generated a mutant in which this helix in Smad2 was replaced with the equivalent region of Smad1 (Smad2C H2

Swap; Chen *et al* 1998). This mutant was inactive, indicating that  $\alpha$ -helix2 of Smad2 is also required for binding to Mixer, Milk and endogenous DEBP (Table 1).

## 5 Identification of a Smad interaction motif

The common property of Smad2 interaction shared by Mixer, Milk and Fast-1 prompted us to analyse sequence similarities between these transcription factors. Whereas Mixer and Milk belong to the same family of homeodomain transcription factors, Fast-1 belongs to an unrelated family of winged-helix/forkhead transcription factors (Chen *et al* 1996; Kaufmann and Knöchel, 1996). We identified a short conserved sequence present in the C-terminal region of Mixer, Milk, and Xenopus Fast-1, which was flanked by sequences of no obvious similarity. It is characterized by a completely conserved PP(T/N)K core, flanked by other highly-conserved residues (Figure 4A; black line above sequences). This sequence is also present in human Fast-1 and mouse Fast-2, which also interact with Smad2 (Labbé *et al* 1998; Zhou *et al* 1998; Liu *et al* 1999); Figure 4A). Significantly, the PP(T/N)K core motif is absent in Mix.1, which does not interact with Smad2.

To address the potential role of this PP(T/N)K-containing sequence in Smad2 interaction, a series of C-terminal deletion mutants of Mixer, Milk and Fast-1 were produced *in vitro* and assayed by bandshift for their ability to bind the DE and interact with GSTSmad2C. Deletion of the PP(T/N)K-containing sequence in the context of either Mixer, Milk or Fast-1 resulted in the loss of interaction with GSTSmad2C, demonstrating that this sequence is necessary for interaction with

Smad2C (Figure 4B). Further C-terminal deletions that impinge on the homeodomains of Mixer or Milk completely abolished DNA binding as expected.

The role of the PP(T/N)K core motif for Smad2 interaction was investigated in more detail, by mutating the two conserved prolines of the PP(T/N)K motif to alanine in the context of full length Mixer [Mixer (PP mut)]. This mutation was sufficient to completely abolish the interaction of Mixer with GSTSmad2C, without affecting its DNA binding properties (Figure 4C). This short motif is thus absolutely required for Smad2 interaction.

It was important to establish that these PP(T/N)K-containing transcription factors could also interact with the effector domain of Smad2 in the absence of DNA. Mixer, Milk, and Fast-1 interacted efficiently with Sepharose-bound Smad2C; but Mixer (PP mut) and Mix-1, the family member that does not contain the PP(T/N)K-containing interaction motif, did not (Figure 4D).

Taken together with the results in the previous sections, we have identified Mixer and Milk as Smad2-interacting proteins, and define the PP(T/N)K-containing sequence present in the C-terminal domain of these homeodomain proteins and also present in XFast-1 and mFast-2 as a Smad Interaction Motif (SIM) essential for Smad2 interaction.

#### **The Smad interaction motif (SIM) is sufficient to bind Smad2**

We next investigated whether the SIM was sufficient to interact with Smad2 by two different assays. First we tested whether a peptide containing 25-amino acids of Mixer incorporating the SIM (residues 283-307; Figure 4A) could compete with

Mixer for binding Smad2C. Indeed, wild type peptide corresponding to approximately 10 and 30-fold molar excess over GSTSmad2C was sufficient to inhibit the interaction of Mixer with GSTSmad2C (Figure 5A, lanes 9,10). The same quantity of the equivalent peptide with the 2 prolines of the PP(T/N)K motif mutated was ineffective (lanes 13,14). This indicates that the peptide alone is sufficient to bind Smad2C, thus preventing full length Mixer binding.

If, as our data above suggests, the same SIM in Fast-1 is used to recruit active Smad2 in the complex ARF, then we would expect that the SIM-containing peptide would be able to disrupt the formation of endogenous ARF. This is exactly what we observe (Figure 5B). Wild type peptide, but not the mutant, is sufficient to inhibit the formation of endogenous *Xenopus* ARF complex (lanes 2-11). Thus the SIM-containing peptide can bind to endogenous Smad2 and inhibit Smad2's interaction with Fast-1.

15

#### **Mixer recruits an active Smad complex *in vivo***

We have shown that Mixer and Milk interact with the C-terminal effector domain of Smad2. However, activated Smad2 exists *in vivo* as a complex with Smad4 (Massagué, 1998). We therefore sought direct evidence using co-immunoprecipitation and bandshift assays that these homeodomain proteins could form stable complexes with ligand-activated Smad2/Smad4 *in vivo*. NIH3T3 cells were used for these experiments since they do not express Mixer or Milk, and this avoided complications of the synthesis of Mix family members in response to activin in *Xenopus* embryo explants. TGF- $\beta$  was used to stimulate the NIH3T3s as it activates Smad2 and Smad4 in the same way as activin (Liu *et al* 1997) and NIH3T3s respond strongly to TGF- $\beta$ , and not to activin.

25

This heterologous system has considerable advantages which allow us to assess the relative importance of Mixer versus the Mixer/Smad complex for transcription via the DE. In particular, 3T3s lack endogenous Mixer/Milk/Bix, but express Smads almost identical to the *Xenopus* Smads, which can be activated in exactly the same way as in a *Xenopus* embryo. This enables us not only to demonstrate that a Mixer/Smad complex forms in response to TGF- $\beta$  but we can show that this complex is ~ 25-times more transcriptionally active than Mixer alone. Moreover, the double point mutant that doesn't interact with Smads, cannot mediate TGF- $\beta$ -induced transcription.

These experiments cannot be easily interpreted when performed in animal caps because the results are complicated by the fact that activin induces Mixer/Milk/Bix expression. For instance, the induction of these endogenous genes makes it very difficult to interpret the effect of any mutant derivative, such as Mixer PP mutant.

Figure 6A shows a co-immunoprecipitation assay in which Flag-tagged, Mixer, Mixer (PP mut) or Fast-1 were immunoprecipitated from cells incubated for 1 hour with or without TGF- $\beta$  and then Western blotted with anti-myc antibody to detect the presence of co-immunoprecipitating myc-tagged Smads. Equal expression of protein was confirmed by Western blotting using anti-Flag or anti-myc antibody of whole cell extracts (Figure 6A, middle and bottom panels). In these conditions of overexpressed Smads, Fast-1 constitutively interacted with both Smad2 and Smad4 (Figure 6A, top panel; but see below). In contrast, in the absence of ligand, Mixer interacted with Smad2 only, but Mixer clearly associated with both Smad2 and Smad4 after TGF- $\beta$  stimulation (Figure 6A, top panel). Mutation of the two prolines in the SIM in Mixer (Mixer PP mut) completely abolished the formation

of this Mixer/Smad complex *in vivo* (Figure 6A, top panel). Thus Mixer can form a ligand-dependent complex with activated Smad2 and Smad4 *in vivo* in the absence of DNA and this requires the integrity of the SIM.

5 We next determined by bandshift assay on a single DE probe whether Mixer could form a stable TGF- $\beta$  inducible complex with endogenous Smads on DNA. As a control for the Smad antibodies we demonstrated that they could supershift the Fast-1/Smad2/Smad4 complex, ARF on the ARE probe (Figure 6B). ARF is strongly ligand-inducible in these conditions (lanes 2,7), and clearly contains Fast-1 and endogenous Smad2 and 4 as shown by antibody supershifts (lanes 1-10; 10 Chen *et al* 1996; Chen *et al* 1997). A Mixer/DNA complex is seen in extracts from cells transfected with Flag-tagged Mixer (Figure 6C, lanes 1-8). In addition, a strong TGF- $\beta$ -induced Mixer-Smad complex was detected with extracts made from cells induced with TGF- $\beta$  for 1 hour (compare lanes 1 and 5). This Mixer-Smad 15 complex contained endogenous Smad2 and 4 as demonstrated by the antibody supershifts (lanes 6-8). We could additionally prove that TGF- $\beta$ -inducible Mixer/Smad complex must contain Mixer as well as the Smads, since no such complex was formed in cells expressing Flag-tagged Mixer (PP mut), which does not interact with Smads (lanes 9-16).

20

### **Mixer and Milk confer TGF- $\beta$ inducibility on the DE**

Having demonstrated that Mixer forms a DNA-binding complex with activated endogenous Smads in response to TGF- $\beta$ , we investigated whether this complex was transcriptionally active. A DE-driven CAT reporter gene was inactive in 25 NIH3T3 cells and did not respond to TGF- $\beta$  induction (Figure 7A). Co-transfection of Smad2 and Smad4 had no effect, indicating that the Smads could

not activate transcription alone. Mixer displayed very little transcriptional activity in the absence of TGF- $\beta$ . However, it could confer very strong TGF- $\beta$ -dependent transcriptional activation on the DE (~ 25-fold induction; Figure 7A). In contrast, the mutant of Mixer that does not bind Smad2 (Mixer PP mut) was completely  
5 inactive (Figure 7A). This provided strong evidence that TGF- $\beta$  induction of transcription via Mixer required recruitment of endogenous Smads. This was corroborated by the observation that overexpression of Smad2 and Smad4 potentiated transcription via Mixer in the absence of TGF- $\beta$  stimulation. Milk also conferred TGF- $\beta$  inducibility on the DE. However, Mix.1 was inactive, consistent  
10 with the fact that it does not interact with Smad2 (Figure 7A). These reporter gene assays were performed with four tandem DE elements. Mixer and Milk were also sufficient to confer TGF- $\beta$  induced transcription onto a single DE, albeit at a lower level (data not shown). TGF- $\beta$  induced transcription mediated by the homeodomain proteins were stronger than that elicited by Fast-1 on the ARE  
15 (Figure 7A; Liu *et al* 1997), which mirrors what we observe in *Xenopus* animal cap assays (Figure 1A).

Given that the TGF- $\beta$  activation of transcription mediated by Mixer results from Mixer's interaction with the Smads, we would expect it to be independent of new  
20 protein synthesis. We show that this is indeed the case using globin reporter system where mRNA levels are measured directly by RNase protection. TGF- $\beta$ -induced transcription via the DE is absolutely dependent on Mixer (Figure 7B, lanes 1,2,5,6) and crucially is not decreased when cycloheximide was added at the same time as TGF- $\beta$  (lanes 5-8). Thus when Mixer is present TGF- $\beta$  induced  
25 transcription does not require on-going protein synthesis.

### Mixer and Milk are expressed appropriately to be endogenous inducers of *goosecoid*

If, as we propose, Mixer and/or Milk are endogenous inducers of *goosecoid* then we would expect them to be expressed in the same domain as *goosecoid*. We investigated the spatial expression patterns of *Mixer*, *Milk* and *goosecoid* by whole mount *in situ* hybridisation in St 10.25 embryos (Figure 8A). In these experiments *Mixer* and *Milk* mRNAs are stained turquoise and *goosecoid* mRNA, deep purple. *Goosecoid* is expressed in the dorsal marginal zone (above the dorsal lip - arrow). *Mixer* and *Milk* are expressed more widely. *Mixer* is expressed throughout the marginal zone (prospective mesoderm) and in vegetal cells (prospective endoderm) and *Milk* is expressed in the dorsal and lateral marginal zone, and in vegetal cells. Mixer was previously thought to be expressed exclusively in the endoderm at stage 12 (Henry and Melton, 1998). It is possible that the expression we see in the mesoderm at Stage 10.25 is lost at later stages. In the double *in situs*, the overlapping purple stain of the *goosecoid* signal and the turquoise from the *Mixer* or *Milk* gives a dark blue stain, seen in the dorsal marginal zone (Figure 8A).

We also addressed the timing of *Mixer* and *Milk* expression in *Xenopus* embryos. *Mixer* and *Milk* are both expressed before the major upregulation of *goosecoid* expression (Figure 8B, lanes 1-8). In addition, the *Milk* RNase protection probe also detects a *Milk-related* transcript, likely to be derived from the highly-related *Bix* genes (see below and Experimental procedures; Ecochard *et al* 1998; Tada *et al* 1998).

Thus the expression patterns of *Mixer* and *Milk* overlap with *goosecoid* in the dorsal marginal zone at early gastrula stages, and *Mixer* and *Milk* are both



expressed before *goosecoid* consistent with them being responsible for induction of *goosecoid*.

### **The role of Mix family members in meso-endoderm induction**

- 5 Finally we addressed the timing of *Mixer* and *Milk* expression relative to the timing of production of the endogenous activin-like signal. In *Xenopus* embryos the major secreted mesoderm-inducing activin-like signal is zygotic and requires the maternal transcription factor VegT for its production (Kimelman and Griffin, 1998; Zhang *et al* 1998). Maternal transcription factors, as well as being  
10 responsible for producing the activin-like signal that gives rise to the active Smad complexes, might also be responsible for the synthesis of the transcription factors the Smads interact with. Inductions by maternal factors such as VegT should be not be abolished by incubating the embryos in cycloheximide prior to Stage 8.
- 15 The expression of *Mixer* was virtually all abolished by the cycloheximide treatment, suggesting that it is solely induced by zygotic activators (Figure 8B lanes 1-16). By contrast, *Milk* and the *Milk-related* gene were strongly activated in untreated embryos, and some of this activation remained in cycloheximide-treated embryos (Figure 8B lanes 1-16). This suggests that these genes are weakly induced  
20 by maternal activators, and their expression is reinforced by zygotic activators. The temporal expression patterns and sensitivity to cycloheximide of the highly related *Bix* genes, *Bix1*, 3 and 4, were identical to *Milk-related* gene in this assay (data not shown). *Milk-related* is most likely to be *Bix3* from the size of the protected fragment in the RNase protection. *Bix3* also contains a well conserved PP(T/N)K-  
25 containing SIM.

Thus in the embryo, Milk and Milk-related are likely to be the earliest endogenous Mix family partner for Smads to initiate transcription of meso-endodermal genes. A zygotic signal, probably the endogenous activin-like signal (Ecochard *et al* 1998; Tada *et al* 1998) induces the synthesis of additional Milk, Milk-related and also Mixer. In fact this is likely to correspond to the DEBP that we detect at Stage 10.5/11. The complexes this Mixer/Milk/Milk-related form with Smads could be responsible for maintaining the activin-induced transcription of meso-endodermal genes (Figure 8D).

## 10 Discussion

### **Mixer and Milk recruit Smads to the *goosecoid* DE to regulate activin/TGF- $\beta$ responsive transcription**

In this example we have investigated the mechanism of activin-responsive transcription via the distal element of the *Xenopus goosecoid* promoter. We have shown that paired-like homeodomain transcription factors of the Mix family, Mixer and Milk, but not Mix.1, mediate activin/TGF- $\beta$ -induced transcription *via* the DE by interacting specifically with the effector domain of Smad2, thereby recruiting active Smad2/Smad4 complexes to this element (Figure 8C). We demonstrate that the molecular basis for the specificity in the Smad2 interaction is the  $\alpha$ -helix 2 of the Smad2 MH2 domain (Shi *et al* 1997). We show that Mixer forms a TGF- $\beta$ -inducible complex with endogenous Smad2 and Smad4 at the *goosecoid* DE within 1 hour of ligand stimulation, and we can demonstrate that the Smads are essential for transcriptional regulation mediated by this complex, since the Mixer/Smad complex is approximately 25-fold more transcriptionally active than Mixer alone.

Our results also reveal that activated Smads are recruited to different promoter elements by a common mechanism. We have identified a short Smad interaction motif (SIM), characterized by the core sequence PP(T/N)K, in the C-terminal region of Mixer and Milk, which is both necessary and sufficient for these proteins

5 to interact with the effector domain of Smad2. Crucially it is also conserved in the C-terminal regions of the winged-helix/forkhead Smad2-interacting proteins: *Xenopus* Fast-1, human Fast-1 and mouse Fast-2 (Chen *et al* 1996; Chen *et al* 1997; Labbé *et al* 1998; Zhou *et al* 1998). This indicates that transcription factors of completely different DNA-binding specificity recruit activated Smads to distinct

10 promoter elements *via* the same protein-protein interaction. This finding now explains why activin-responsive elements in the promoters of different *Xenopus* genes share so little sequence similarity (Howell and Hill, 1997).

#### Activation of transcription by Smad/transcription factor complexes

15

The Smads appear to require other transcription factors to recruit them to DNA because they interact with DNA themselves either very weakly (Smad3 and Smad4) or not at all (Smad2; (Shi *et al* 1998; Hill, 1999). There appears to be a broad range of transcription factor/Smad interactions. At one extreme are the

20 ~~functionally-cooperative interactions such as that seen between the *Drosophila* homeodomain protein, tinman and MAD and MEDEA (Xu *et al* 1998) where no~~

physical contact between the transcription factor and Smads has been reported. At the other extreme is the direct transcription factor-Smad complexes such as that described here and for Fast-1 and also AP-1 family members (Derynck *et al* 1998).

25 As well as forming transcriptionally active complexes, activated Smads may also be able to release repressors from DNA. Recent work suggests that the

homeodomain protein Hoxc-8 functions as a repressor, and interaction with activated Smad1 releases Hoxc-8 from DNA (Shi *et al* 1999).

The interaction of Smads with distinct transcription factors must contribute to cell-type specificity of TGF- $\beta$  responses, allowing specific genes to be up-regulated only in cells where the essential co-operating transcription factor is also expressed. This is likely to be of particular significance in the patterning of the early *Xenopus* embryo. The same signalling pathway will activate different genes in distinct regions of the embryo depending on the particular Smad-recruiting transcription factors expressed by the cells in that region. In addition, differential affinities of specific transcription factors for Smads, coupled with the presence or absence of Smad binding sites on adjacent DNA could allow distinct genes to be activated by different levels of active Smad complexes. This sort of mechanism might underlie the morphogenetic properties of TGF- $\beta$  family members, whereby different doses of TGF- $\beta$  ligands elicit different transcriptional responses (Green and Smith, 1990). Determination of the relative affinities of Mixer, Milk, Bix proteins and Fast-1 for Smad complexes will be important to test these ideas.

### **Regulation of *goosecoid***

Previous studies of *Xenopus goosecoid* regulation indicated that there was an activity in the vegetal hemisphere and marginal zone which could regulate transcription via the DE (Watabe *et al* 1995). This was not sufficient for regulation of the *goosecoid* promoter, which additionally required regulation through the PE by a Wnt-induced transcription factor. The combination of these activities confined *goosecoid* expression to the dorsal marginal zone of the early gastrula (Watabe *et al* 1995; Laurent *et al* 1997). We now propose that the activin-induced DE

transcriptional activity corresponds to a SIM-containing member(s) of the Mix family, complexed with activated Smad2/Smad4. Our experiments do not at present allow us to distinguish between Mixer, Milk or other Bix-family members as the endogenous transcription factor responsible. In addition a Fast-1/Smad complex may also be involved in the context of the whole *goosecoid* promoter, since a functional Fast-1 binding site was identified in the mouse *goosecoid* promoter, which is largely conserved downstream of the DE in the *Xenopus* promoter (Labbé *et al* 1998). However, this Fast-1 binding site in the mouse promoter is not sufficient for efficient TGF- $\beta$ /activin induced transcription, and requires adjacent Smad4 binding sites (Labbé *et al* 1998), which are not conserved in the *Xenopus* promoter (Watabe *et al* 1995). It will be important for the future to investigate possible functional interactions between Fast-1, Mix family members and activated Smads on the *Xenopus goosecoid* promoter.

## 15    **The role of Mixer and Milk in meso-endodermal induction in *Xenopus* embryos**

Previous work had already implicated Mixer and Milk/Bix in endodermal and mesodermal differentiation, based on experiments in which they were overexpressed in prospective ectoderm (animal caps) (Ecochard *et al* 1998; Henry and Melton, 1998; Tada *et al* 1998). However the underlying mechanism was unknown. Our data suggest that Mixer/Milk/Bix have little inherent transcriptional activity, but require bound Smads activated by an endogenous activin-like signal to increase their transcriptional potential and thus activate meso-endodermal genes. We would therefore predict that the family member Mix.1, which does not interact with Smads, would have a different activity *in vivo*. Indeed, in contrast to Mixer

and Milk, Mix.1 does not induce endoderm when overexpressed in animal caps (Ecochard *et al* 1998; Henry and Melton, 1998; Tada *et al* 1998).

Our interaction data, together with the expression patterns of these homeodomain proteins allows us to propose a model for meso-endodermal formation in the *Xenopus* embryo (Figure 8D). The major activin-like meso-endoderm-inducing activity that would activate Smad2 and Smad4 is zygotic, and requires the maternal transcription factor, VegT for its production (Kimelman and Griffin, 1998; Zhang *et al* 1998). A good candidate for this ligand is the Vg-1-related protein, *derrière* (Sun *et al* 1999). Our experiments indicated that Milk and Milk-related, which is probably Bix3, are also induced (weakly) in *Xenopus* embryos by a maternal activator (Figure 8D). This could be VegT itself, since the Bix genes have been shown to be VegT targets (Tada *et al* 1998). Thus low levels of Milk and Milk-related would be available to bind the Smad2/4 complexes activated by the zygotic activin-like ligand to initiate transcription of downstream genes like goosecoid (Figure 8D). In addition, there may be low levels of ubiquitously maternally expressed Milk/Bix genes that would account for the cycloheximide-insensitive activin-induced transcription of the DE seen in the animal caps in Figure 1A. Milk and Milk-related and also Mixer are themselves induced by the zygotic activin-like signaling pathway (Ecochard *et al* 1998; Henry and Melton, 1998; Tada *et al* 1998). We propose that these proteins would be involved in maintaining transcription in response to the zygotic activin-like ligand through their formation of transcriptionally active complexes with activated Smads (Figure 8D).

In conclusion, our data establish members of the Mix family as transcriptional partners for Smads, responsible for mediating activin/TGF- $\beta$  responsive

transcription in the *Xenopus* embryo via paired-like homeodomain binding sites. It is intriguing that there are a number of highly related genes in the family with apparently identical DNA-binding specificity (Ecochard *et al* 1998; Henry and Melton, 1998; Tada *et al* 1998), and similar expression patterns. Understanding  
 5 how they each contribute to the patterning of the *Xenopus* embryo will be an important task for the future.

## Experimental procedures

### Plasmid constructs

10 Mix.1 (Rosa, 1989), Milk (Ecochard *et al* 1998) and Mixer (Henry and Melton, 1998) were isolated by PCR from a Stage 11 *Xenopus* cDNA library and their coding sequences and that of Fast-1 (Chen *et al* 1997) were subcloned into pFTX5 (Howell and Hill, 1997) and EF-Flag. Human Smad4 and *Xenopus* Smad2 were subcloned in EF-Myc. EF-Flag and EF-myc were derivatives of EF-Plink (Hill *et al* 1995). Prolines 290 and 291 of Mixer were mutated to alanines by PCR. In  
 15 DE<sub>4</sub>-CAT and ARE<sub>3</sub>-CAT four copies of the *goosecoid* DE or three copies of the Mix.2 ARE are upstream of the minimal  $\gamma$ -actin promoter driving CAT. In the globin versions, human  $\beta$ -globin replaced CAT (Howell and Hill, 1997). REF-globin was as described (Howell and Hill, 1997). In GSTSmad2C amino acids  
 20 198-467 of XSmad2 and in GSTSmad1C amino acids 172-468 of XSmad1 were subcloned into pGEX-KG (Poon *et al* 1993). 5' and 3' deletions of GSTSmad2C were made using standard methods and named according to the positions of the deletion such that GSTSmad2C(198–245) lacks sequence following codon 245, whilst GSTSmad2C ( $\Delta$ 207-245) lacks sequence between codons 208 and 244.  
 25 Helix 2 of Smad2 was replaced by Helix 2 of Smad1 in GSTSmad2C using PCR. All constructs were verified by sequencing.

### Oligonucleotides

1. CTAGCCATTAATCAGATTAACGGTGAGCAATTAGA (DE-top),
2. CCGACTAGTATCTGCTGCCCTAAAATGTGTATTCCATGGAAATG (ARE  
5 top),
3. CCGGCTAGCTAGGGAGAGAAGGGCAGACATTTCCATGGAATAC (ARE  
bot),
4. CTAGCCAGTCAGCAGCTGACCGGTGAGCAATTAGA (DE m1 top),
5. CTAGCCAGTCATCAGAGTCACGGTGAGCAAGTCGA (DE m2 top),
- 10 6. CTAGCCATTAATCAGATTAACCTTGTAGCAAGTAGA (DE m3 top),

### GST fusion protein purification, GST “pull-downs” and *in vitro* transcription/translation

Expression of GST-fusion proteins, SDS-PAGE and *in vitro* coupled  
15 transcription/translation in reticulocyte lysate (Promega) were performed using  
standard methods. Mixer, Milk and Fast-1 C-terminal deletion mutants were  
synthesized *in vitro* using linear templates generated by restriction enzyme  
digestion. For “pull-down” experiments, [<sup>35</sup>S]-labelled transcription factors were  
mixed with GST- or GST-Smad2C-Sepharose beads for 2h at 4°C in 20 mM Tris  
20 pH7.5, 20% glycerol, 1 mM-EDTA, 5 mM MgCl<sub>2</sub>, 0.1% NP40 and 220 mM-NaCl.

The beads were washed three times with five bead volumes of binding buffer, and  
the protein remaining bound to the beads were analysed by SDS-PAGE followed  
by autoradiography.

### 25 Embryo manipulations, RNase protection assays and *in situ* hybridizations



The production, maintenance and manipulation of *Xenopus* embryos was previously described (Howell and Hill, 1997). mRNA for microinjection was generated *in vitro* (Howell and Hill, 1997) and injected at the 1-cell stage; 200 pg Activin  $\beta$ A mRNA per embryo, 1.5 ng mRNA encoding myc-tagged Mix.1, Mixer, or Milk. When embryos were treated with cycloheximide, it was added at 20  $\mu$ g/ml in 0.1X NAM 30 min before St 8. RNA isolation and RNase protection assays were performed as described (Howell and Hill, 1997). The antisense probes were as follows: human  $\beta$ -globin (Howell and Hill, 1997); *goosecoid* (Blumberg *et al* 1991); *Xenopus FGF receptor*, protecting amino acids 539-580; *Mixer*, amino acids 173-237; *Milk*, amino acids 143-226. The *Milk* probe also detects a smaller product, whose size and expression characteristics are consistent with it being the protected fragment of the highly Milk-related gene *Bix3* (Tada *et al* 1998). Whole mount *in situ* hybridizations were carried out essentially as described (Harland, 1991), using probes against *goosecoid*, *Mixer* and *Milk* which were identical to those used in the RNase protections, either singly or in combination.

### Transfections

NIH3T3 cells were transfected using lipofectamine (Gibco BRL). The following amounts of plasmids were used per 6-cm dish for transcriptional assays, 0.5  $\mu$ g CAT-reporters, 0.2  $\mu$ g of transcription factors, 0.3  $\mu$ g of EF-Smad2 and EF-Smad4 and 0.5  $\mu$ g EF-LacZ as an internal control for transfection efficiency, as indicated in the Figure legends. For globin transcriptional assays, 1  $\mu$ g globin reporter, 0.45  $\mu$ g REF-globin and 0.2  $\mu$ g Mixer were transfected. For immunoprecipitations, two 6 cm plates were transfected with 0.6  $\mu$ g of each plasmid. For the bandshift assay, one 6-cm plate was transfected with 1.2  $\mu$ g transcription factor. The amounts of DNA transfected was kept constant by adding control plasmid EF-plink as

appropriate. Following transfection, cells were maintained 18 hr in DMEM containing 10% FBS, before induction by TGF- $\beta$ 1 (2 ng/ml, Calbiochem) for times indicated in the Figure legends.

## 5 Transcriptional assays

After induction, cells were lysed in 200  $\mu$ l of 20 mM Tris-HCl pH7.5, 150 mM NaCl, 1 mM EDTA and 0.5 % NP40. CAT assays were performed exactly as previously described (Hill *et al* 1993).  $\beta$ -galactosidase assays were performed using CDGP (Calbiochem) as a substrate and quantitated spectrophotometrically.

10 RNA was extracted from NIH3T3 cells for the globin assays as described (Hill *et al* 1994) and the RNase protection assays were as above.

## Immunoprecipitation

After induction, cells were lysed in 100  $\mu$ l buffer containing 20 mM Tris HCl pH 7.4, 150 mM NaCl, 1 mM EDTA, 1 mM EGTA, 5 mM NaF, 10 mM  $\beta$ -glycerophosphate, 10% glycerol, 1% Triton and protease inhibitors: 10  $\mu$ g/ml Leupeptin, E-64, Aprotinin, 20  $\mu$ g/ml Pepstatin, 0.5 mM Benzamidine and 0.4 mM Pefabloc SC. Flag-tagged transcription factors were immunoprecipitated with anti-Flag M2 affinity gel (Sigma) for 2 hr at 4°C, and washed three times with lysis  
20 buffer. Immunoprecipitates were separated by SDS 15% polyacrylamide gel electrophoresis and Western-blotted with anti-myc antibody 9E10.

## Bandshift assays and peptides

Bandshift probes corresponding to the ARE (oligonucleotides 2 and 3) and DE  
25 (oligonucleotide 1 and its complement) were labelled with [ $\alpha^{32}$ P]dATP and [ $\alpha^{32}$ P]dCTP by PCR. Competitions were performed with double-stranded

oligonucleotides: DE m1, DE m2, DE m3 or ARE (produced by annealing and filling in oligonucleotides 2 and 3). Whole cell *Xenopus* embryo extracts were prepared by homogenizing embryos in buffer (10  $\mu$ l per embryo) containing 200 mM KCl, 50 mM Tris-HCl pH 7.4, 10% glycerol, 25 mM  $\beta$ -glycerophosphate, 1 mM EGTA, 1 mM EDTA, 2 mM DTT, and protease inhibitors as above. Lysates were cleared by repeated centrifugation. Binding reactions were performed with 30  $\mu$ g of protein extract incubated with 0.2 ng DE probe in 20  $\mu$ l of buffer containing 140 mM KCl, 8 mM  $MgCl_2$ , 12.5 mM  $\beta$ -glycerophosphate, 1 mM EGTA, 1 mM EDTA, 1 mM NaF and 0.5  $\mu$ g poly(dI-dC), 2 mM DTT, and protease inhibitors for 20 min at room temperature. 20 ng of purified GST-fusion proteins were used to study interactions with GSTSmad2C. Extracts for the ARF bandshift in Figure 6B were prepared from activin-injected Stage 8 embryos and the bandshift conditions was as described (Huang *et al* 1995). For *in vitro* translated Mix family members, bandshift conditions were as described (Wilson *et al* 1993). For *in vitro* translated Fast-1, the final buffer concentrations were 8 mM Hepes pH 7.6, 90 mM KCl, 5 mM  $MgCl_2$ , 4 mM  $\beta$ -glycerophosphate, 40  $\mu$ M EDTA, 40  $\mu$ M Spermidine, 2  $\mu$ g poly(dI-dC), 5% glycerol. Extracts from NIH3T3 cells transfected with Flag-tagged transcription factors were prepared as described (Marais *et al* 1993) and final bandshift conditions were: 14  $\mu$ g total protein in 10 mM Hepes pH 7.5, 15% glycerol, 210 mM KCl, 5.5 mM  $MgCl_2$ , 0.2% Triton, 5 mM EGTA, 2.5 mM EDTA, 2  $\mu$ g poly(dI-dC), 0.25 mM DTT and protease inhibitors with 0.2 ng labelled probe. Specific antibodies (1  $\mu$ l) anti-Smad2 (Nakao *et al* 1997), anti-Smad4 (B8; Santa Cruz) or anti-Flag were added to the binding reactions. In all cases electrophoresis was in 5% polyacrylamide gels/0.5X TBE containing 2.5% glycerol.

The wild type SIM-containing peptide used in Figure 5 was:

Biotin.Amino hexanoic acid-

RQIKIWFQNRRMKWKKLLMDFNNFPPNKTITPDMNVRIPPI. The first 16 amino acids are from the helix 3 of *Antennapedia* which allows internalization of these peptides into live cells (Derossi *et al* 1998); the last 25 amino acids are codons 283-307 of Mixer. The mutant was the same except that the 2 prolines at positions 26 and 27 were alanines. Peptides were included in the binding reactions at the concentrations given in the Figure legend.

human recombinant activin A was supplied by NHPP (lot 15365-36(1)).

## References

Blumberg, B., Wright, C. V., De Robertis, E. M., and Cho, K. W. (1991). Organizer-specific homeobox genes in *Xenopus laevis* embryos. *Science* 253, 194-196.

Candia, A. F., Watabe, T., Hawley, S. H., Onichtchouk, D., Zhang, Y., Derynck, R., Niehrs, C., and Cho, K. W. (1997). Cellular interpretation of multiple TGF- $\beta$  signals: intracellular antagonism between activin/BVg1 and BMP-2/4 signaling mediated by Smads. *Development* 124, 4467-4480.

Chen, X., Rubock, M. J., and Whitman, M. (1996). A transcriptional partner for MAD proteins in TGF- $\beta$  signalling. *Nature* 383, 691-696.

Chen, X., Weisberg, E., Fridmacher, V., Watanabe, M., Naco, G., and Whitman, M. (1997). Smad4 and FAST-1 in the assembly of activin-responsive factor. *Nature* 389, 85-89.

- Chen, Y. G., Hata, A., Lo, R. S., Wotton, D., Shi, Y., Pavletich, N., and Massagué, J. (1998). Determinants of specificity in TGF- $\beta$  signal transduction. *Genes Dev.* *12*, 2144-2152.
- Derossi, D., Chassaing, G., and Prochiantz, A. (1998). Trojan peptides: the  
5 penetratin system for intracellular delivery. *Trends Cell Biol* *8*, 84-87.
- Derynck, R., Zhang, Y., and Feng, X. H. (1998). Smads: transcriptional activators of TGF- $\beta$  responses. *Cell* *95*, 737-740.
- Ecochard, V., Cayrol, C., Rey, S., Foulquier, F., Caillol, D., Lemaire, P., and Duprat, A. M. (1998). A novel *Xenopus* mix-like gene milk involved in the control  
10 of the endomesodermal fates. *Development* *125*, 2577-2585.
- Funahashi, J., Sekido, R., Murai, K., Kamachi, Y., and Kondoh, H. (1993).  $\delta$ -crystallin enhancer binding protein  $\delta$ EF1 is a zinc finger-homeodomain protein implicated in postgastrulation embryogenesis. *Development* *119*, 433-446.
- Green, J. B., and Smith, J. C. (1990). Graded changes in dose of a *Xenopus* activin  
15 A homologue elicit stepwise transitions in embryonic cell fate. *Nature* *347*, 391-394.
- Gurdon, J. B., Harger, P., Mitchell, A., and Lemaire, P. (1994). Activin signalling and response to a morphogen gradient. *Nature* *371*, 487-492.
- Harland, R., and Gerhart, J. (1997). Formation and function of Spemann's  
20 organizer. *Ann. Rev. Cell Dev. Biol.* *13*, 611-667.
- 
- Harland, R. M. (1991). In situ hybridization: an improved whole-mount method for *Xenopus* embryos. *Methods Cell Biol.* *36*, 685-695.
- 
- Heasman, J. (1997). Patterning the *Xenopus* blastula. *Development* *124*, 4179-4191.
- 25 Henry, G. L., Brivanlou, I. H., Kessler, D. S., Hemmati-Brivanlou, A., and Melton, D. A. (1996). TGF- $\beta$  signals and a pre-pattern in *Xenopus laevis* endodermal development. *Development* *122*, 1007-1015.
-

- Henry, G. L., and Melton, D. A. (1998). Mixer, a homeobox gene required for endoderm development. *Science* 281, 91-96.
- Hill, C. S. (1999). The Smads. *Int. J. Biochem. Cell Biol.* in press.
- Hill, C. S., Marais, R., John, S., Wynne, J., Dalton, S., and Treisman, R. (1993).  
 5 Functional analysis of a growth factor-responsive transcription factor complex. *Cell* 73, 395-406.
- Hill, C. S., Wynne, J., and Treisman, R. (1994). Serum-regulated transcription by serum response factor (SRF): a novel role for the DNA-binding domain. *EMBO J.* 13, 5421-5432.
- 10 Hill, C. S., Wynne, J., and Treisman, R. (1995). The Rho family GTPases RhoA, Rac1, and CDC42Hs regulate transcriptional activation by SRF. *Cell* 81, 1159-1170.
- Hogan, B. L. (1996). Bone morphogenetic proteins in development. *Curr. Opin. Genet. Dev.* 6, 432-438.
- 15 Hogan, B. L. M., Blessing, M., Winnier, G. E., Suzuki, N., and Jones, C. M. (1994). Growth factors in development: the role of TGF- $\beta$  related polypeptide signalling molecules in embryogenesis. *Development*, Suppl. 53-60.
- Howell, M., and Hill, C. S. (1997). XSmad2 directly activates the activin-inducible, dorsal mesoderm gene XFKH1 in *Xenopus* embryos. *EMBO J.* 16,  
 20 7411-7421.
- 
- Howell, M., Itoh, F., Pierreux, C. E., Valgeirsdottir, S., Itoh, S., ten Dijke, P., and Hill, C. S. (1999). *Xenopus* Smad4 $\beta$  is the co-Smad component of developmentally-regulated transcription factor complexes responsible for induction of early mesodermal genes. *Dev. Biol.* in press.
- 25 Huang, H.-C., Murtaugh, L. C., Vize, P. D., and Whitman, M. (1995). Identification of a potential regulator of early transcriptional responses to mesoderm inducers in the frog embryo. *EMBO J.* 14, 5965-5973.
-

- Kaufmann, E., and Knöchel, W. (1996). Five years on the wings of forkhead. *Mech. Dev.* 57, 3-20.
- Kimelman, D., and Griffin, K. J. (1998). Mesoderm induction: a postmodern view. *Cell* 94, 419-421.
- 5 Kingsley, D. M. (1994). The TGF- $\beta$  superfamily: new members, new receptors, and new genetic tests of function in different organisms. *Genes Dev.* 8, 133-146.
- Kispert, A., Koschorz, B., and Herrmann, B. G. (1995). The T protein encoded by Brachyury is a tissue-specific transcription factor. *EMBO J.* 14, 4763-4772.
- Labbé, E., Silvestri, C., Hoodless, P. A., Wrana, J. L., and Attisano, L. (1998).
- 10 Smad2 and Smad3 positively and negatively regulate TGF  $\beta$ -dependent transcription through the forkhead DNA-binding protein FAST2. *Mol. Cell* 2, 109-120.
- Laurent, M. N., Blitz, I. L., Hashimoto, C., Rothbacher, U., and Cho, K. W. (1997). The *Xenopus* homeobox gene *twin* mediates Wnt induction of goosecoid
- 15 in establishment of Spemann's organizer. *Development* 124, 4905-4916.
- Liu, B., Dou, C. L., Prabhu, L., and Lai, E. (1999). FAST-2 is a mammalian winged-helix protein which mediates transforming growth factor  $\beta$  signals. *Mol. Cell Biol.* 19, 424-430.
- Liu, F., Pouponnot, C., and Massagué, J. (1997). Dual role of the Smad4/DPC4
- 20 tumor suppressor in TGF $\beta$ -inducible transcriptional complexes. *Genes Dev.* 11, 3157-3167.
- 
- Marais, R., Wynne, J., and Treisman, R. (1993). The SRF accessory protein Elk-1 contains a growth factor-regulated transcriptional activation domain. *Cell* 73, 381-393.
- 25 Massagué, J. (1998). TGF- $\beta$  signal transduction. *Ann. Rev. Biochem.* 67, 753-791.
-

- McKendry, R., Harland, R. M., and Stachel, S. E. (1998). Activin-induced factors maintain goosecoid transcription through a paired homeodomain binding site. *Dev. Biol.* 204, 172-186.
- Nakao, A., Imamura, T., Souchelnytskyi, S., Kawabata, M., Ishisaki, A., Oeda, E.,  
 5 Tamaki, K., Hanai, J., Heldin, C. H., Miyazono, K., and ten Dijke, P. (1997). TGF- $\beta$  receptor-mediated signalling through Smad2, Smad3 and Smad4. *EMBO J.* 16, 5353-5362.
- Poon, R. Y., Yamashita, K., Adamczewski, J. P., Hunt, T., and Shuttleworth, J. (1993). The cdc2-related protein p40MO15 is the catalytic subunit of a protein  
 10 kinase that can activate p33cdk2 and p34cdc2. *EMBO J.* 12, 3123-3132.
- Rosa, F. M. (1989). Mix.1, a homeobox mRNA inducible by mesoderm inducers, is expressed mostly in the presumptive endodermal cells of *Xenopus* embryos. *Cell* 57, 965-974.
- Shi, X., Yang, X., Chen, D., Chang, Z., and Cao, X. (1999). Smad1 interacts with  
 15 homeobox DNA-binding proteins in bone morphogenetic protein signaling. *J. Biol. Chem.* 274, 13711-13717.
- Shi, Y., Hata, A., Lo, R. S., Massagué, J., and Pavletich, N. P. (1997). A structural basis for mutational inactivation by the tumour suppressor Smad4. *Nature* 388, 87-93.
- 20 Shi, Y., Wang, Y. F., Jayaraman, L., Yang, H., Massagué, J., and Pavletich, N. P. (1998). Crystal structure of a Smad MH1 domain bound to DNA: insights on DNA binding in TGF- $\beta$  signaling. *Cell* 94, 585-594.
- 
- Sun, B. I., Bush, S. M., Collins-Racie, L. A., LaVallie, E. R., DiBlasio-Smith, E. A., Wolfman, N. M., McCoy, J. M., and Sive, H. L. (1999). derrière: a TGF- $\beta$   
 25 family member required for posterior development in *Xenopus*. *Development* 126, 1467-1482.
-



- Tada, M., Casey, E. S., Fairclough, L., and Smith, J. C. (1998). Bix1, a direct target of *Xenopus* T-box genes, causes formation of ventral mesoderm and endoderm. *Development* 125, 3997-4006.
- Vize, P. D. (1996). DNA sequences mediating the transcriptional response of the *Mix.2* homeobox gene to mesoderm induction. *Dev. Biol.* 177, 226-231.
- 5 Watabe, T., Kim, S., Candia, A., Rothbacher, U., Hashimoto, C., Inoue, K., and Cho, K. W. (1995). Molecular mechanisms of Spemann's organizer formation: conserved growth factor synergy between *Xenopus* and mouse. *Genes Dev.* 9, 3038-3050.
- 10 Whitman, M. (1998). Smads and early developmental signaling by the TGF $\beta$  superfamily. *Genes Dev.* 12, 2445-2462.
- Wilson, D., Sheng, G., Lecuit, T., Dostatni, N., and Desplan, C. (1993). Cooperative dimerization of paired class homeo domains on DNA. *Genes Dev.* 7, 2120-2134.
- 15 Xu, X., Yin, Z., Hudson, J. B., Ferguson, E. L., and Frasch, M. (1998). Smad proteins act in combination with synergistic and antagonistic regulators to target Dpp responses to the *Drosophila* mesoderm. *Genes Dev.* 12, 2354-2370.
- Zhang, J., Houston, D. W., King, M. L., Payne, C., Wylie, C., and Heasman, J. (1998). The role of maternal VegT in establishing the primary germ layers in
- 20 *Xenopus* embryos. *Cell* 94, 515-524.
- ~~Zhou, S., Zawel, L., Lengauer, C., Kinzler, K. W., and Vogelstein, B. (1998). Characterization of human FAST-1, a TGF  $\beta$  and activin signal transducer. *Mol. Cell* 2, 121-127.~~

**Table 1. Mapping the transcription factor interaction domain in Smad2**

5	GST-fusions	endogenous DEBP	Mixer	Milk	Fast-1
	GST	-	-	-	-
10	<b>a</b> Smad2C (198-467)	+	+	+	+
	<b>b</b> Smad2C (198-463)	+	+	+	+
	Smad2C (198-445)	+	+	-	-
	Smad2C (198-440)	+	+	-	-
15	Smad2C (198-426)	+	+	-	-
	Smad2C (198-401)	-	-	-	-
	Smad2C (198-373)	-	-	-	-
	Smad2C (198-345)	-	-	-	-
	Smad2C (198-315)	-	-	-	-
20	Smad2C (198-276)	-	-	-	-
	Smad2C (198-245)	-	-	-	-
	Smad2C ( $\Delta$ 207-245)	+	+	+	+
	Smad2C ( $\Delta$ 207-259)	+	+	+	+
25	<b>c</b> Smad2C ( $\Delta$ 207-268)	+	+	+	+
	<b>c</b> Smad2C ( $\Delta$ 207-321)	-	-	-	-
	Smad1C	-	-	-	-
	Smad2C(H2 swap)	-	-	-	-

30

Interactions with purified GST fusion proteins were detected by bandshift assay using radiolabelled DE or ARE probes as appropriate. DEBP was derived from whole cell extracts of St 10.5 embryos overexpressing activin, and Mixer, Milk and Fast-1 were produced *in vitro*.

35 a The Smad2C protein corresponds to residues 198-467 of Smad2.

b The C-terminal phosphorylation sites (S-465 and S-467) are deleted in this mutant.

c The MH2 domain begins at amino acid W-274.

CLAIMS

1. A polypeptide capable of interacting with a Smad polypeptide wherein the interacting polypeptide comprises the amino acid sequence PP(T/N)K and is less  
5 than 32 amino acids in length.
2. A polypeptide comprising the amino acid sequence PP(T/N)K that is less than 32 amino acids in length.
- 10 3. A polypeptide capable of interacting with a Smad polypeptide wherein the interacting polypeptide comprises the amino acid sequence PP(T/N)K and is not full-length *Xenopus* or human FAST1 or a fragment thereof, mouse FAST2, *Xenopus* Milk, *Xenopus* Mixer, *Xenopus* Bix3, Bix2 or Bix1.
- 15 4. The polypeptide of claim 1 or 3 wherein the Smad polypeptide is Smad2 or Smad3.
5. The polypeptide of any one of claims 1 to 4 wherein the polypeptide is a transcription factor or a fragment thereof.
- 20 6. The polypeptide of any one of claims 3 to 5 wherein the polypeptide is less than 100 amino acids in length.
7. The polypeptide of any of the preceding claims wherein the polypeptide is  
25 between 4 and about 30 amino acids in length.

8. The polypeptide of any of the preceding claims wherein an acidic amino acid residue is present at a position from 3 to 10 residues C-terminal of the amino acid sequence PP(T/N)K and/or a proline residue is present at a position from 5 to 20 residues C-terminal of the amino acid sequence PP(T/N)K.

5

9. The polypeptide of any of the preceding claims comprising the amino acid sequence PPNKTITPDMNVRIPPI or PPNKTITPDMNTIIPQI or PPNKSVFDVLTSHPGD or PPNKSIYDVWVSHPRD or PPTKTITANMNTIIPQM or PPNKSIYDVWVSHPRD or

10 PPNKTVFDIPVYTGHPG or PPNKTITPDMNTIIPQI or LLMDFNFPNKTITPDMNVRIPPI or HSNLMMDFPNKTITPDMNTIIPQI or LDNMLRAMPPNKSVDVLTSHPGD or LDSLFQGVPPNKSVDVWVSHPRD or LMMDISNFPPTKTITANMNTIIPQM or

15 LDALFQGVPPNKSVDVWVSHPRD or LKNAPSDFPNKTVFDIPVYTGHPG or HSNLVMEFPNKTITPDMNTIIPQI.

10. A polypeptide consisting of the amino acid sequence

20 PPNKTITPDMNVRIPPI or PPNKTITPDMNTIIPQI or PPNKSVFDVLTSHPGD or PPNKSIYDVWVSHPRD or PPTKTITANMNTIIPQM or PPNKSIYDVWVSHPRD or PPNKTVFDIPVYTGHPG or PPNKTITPDMNTIIPQI or LLMDFNFPNKTITPDMNVRIPPI or HSNLMMDFPNKTITPDMNTIIPQI or

25 or LDNMLRAMPPNKSVDVLTSHPGD or LDSLFQGVPPNKSVDVWVSHPRD or LMMDISNFPPTKTITANMNTIIPQM or

LDALFQGVPPNKSIIYDVVSHPRD

or

LKNAPSDFFPPNKTVFDIPVYTGHPG

or

HSNLVMEFPPNKTITPDMNTIIPQI.

- 5 11. The polypeptide of any of the preceding claims comprising the amino acid sequence of residues 283 to 307 of Mixer.

12. The polypeptide of any of the preceding claims wherein the said polypeptide is a peptidomimetic compound.

10

13. A molecule comprising a polypeptide as defined in any of Claims 1 to 12 and a further portion, wherein the said molecule is not full-length *Xenopus* or human FAST1 or a fragment thereof, mouse FAST2, *Xenopus* Milk, *Xenopus* Mixer or *Xenopus* Bix2.

15

14. A molecule according to claim 13 wherein the molecule is Biotin.Amino hexanoic acid-

RQIKIWFQNRRMKWKKLLMDFNNFPPNKTITPDMNVRIPPI

- 20 15. A nucleic acid encoding or capable of expressing a polypeptide or molecule according to any one of claims 1 to 14.

16. A nucleic acid complementary to a nucleic acid encoding a polypeptide according to any one of claims 1 to 11.

25

17. An antibody capable of reacting with a polypeptide according to any one of claims 1 to 12.

18. A method of identifying a polypeptide that is capable of interacting with a Smad polypeptide, comprising examining the sequence of a polypeptide and determining that the polypeptide comprises the amino acid sequence PP(T/N)K.

5

19. The method of claim 18 further comprising determining that an acid amino acid residue is present at a position from 3 to 10 residues C-terminal of the amino acid sequence PP(T/N)K and/or a proline residue is present at a position from 5 to 20 residues C-terminal of the amino acid sequence PP(T/N)K.

10 —

20. A method of identifying a compound capable of disrupting or preventing the interaction between a Smad polypeptide and a target polypeptide that is (1) a transcription factor capable of interacting with the said Smad polypeptide and/or (2) a polypeptide capable of interacting with the said Smad polypeptide, the interaction requiring  $\alpha$ -helix2 of the said Smad polypeptide or (3) a polypeptide comprising the amino acid sequence PP(T/N)K, the method comprising measuring the ability of the compound to disrupt or prevent the interaction between the Smad polypeptide and a polypeptide according to any one of claims 1 to 10.

20

21. A compound identified by or identifiable by the method of claim 20.

---

22. A kit of parts comprising a Smad polypeptide and a polypeptide or molecule according to any one of claims 1 to 14.

25

23. A method of disrupting or preventing the interaction between a Smad polypeptide and a target polypeptide that is (1) a transcription factor capable of

---

interacting with the said Smad polypeptide and/or (2) a polypeptide capable of interacting with the said Smad polypeptide, the interaction requiring  $\alpha$ -helix2 of the said Smad polypeptide, the method comprising exposing the Smad polypeptide to a polypeptide or molecule according to any one of claims 1 to 14  
5 or to an antibody according to claim 17 or to a compound according to claim 21.

24. A method of disrupting or preventing the interaction between a Smad polypeptide and a polypeptide comprising the amino acid sequence PP(T/N)K wherein the Smad polypeptide is exposed to a polypeptide or molecule according  
10 to any one of claims 1 to 14 or to an antibody according to claim 17 or to a compound according to claim 21.

25. The method of claim 23 or 24 wherein the Smad polypeptide is Smad2 or Smad3.  
15

26. A compound according to claim 21 or polypeptide or molecule according to any one of claims 1 to 14 or nucleic acid according to claim 15 or 16 or antibody according to claim 17 for use in medicine.

20 27. A method of modulating activin or TGF $\beta$  signalling in a cell *in vitro* wherein the cell is exposed to a polypeptide, molecule, compound, nucleic acid or antibody as defined in claim 26.

---

28. A method of modulating activin or TGF $\beta$  signalling in a cell *in vivo* wherein  
25 the cell is exposed is exposed to a polypeptide, molecule, compound, nucleic acid or antibody as defined in claim 26.

---

29. The method of claim 27 or 28 wherein the cell is a late stage tumour cell.

30. The use of a polypeptide, molecule, compound, nucleic acid or antibody as defined in claim 26 in the manufacture of a medicament for treatment of a patient in need of modulation of activin or TGF $\beta$  signalling.

31. The use of a polypeptide, molecule, compound, nucleic acid or antibody as defined in claim 26 in the manufacture of a medicament for treatment of a patient with cancer.

32. The use of a polypeptide, molecule, compound, nucleic acid or antibody as defined in claim 26 in the manufacture of a medicament for treatment of a patient in need of reducing extracellular matrix deposition, encouraging tissue repair and/or regeneration, tissue remodelling or healing of a wound, injury or surgery, or reducing scar tissue formation arising from injury to the brain.

33. The use of a polypeptide, molecule, compound, nucleic acid or antibody as defined in claim 26 in the manufacture of a medicament for treatment of a patient with or at risk of end-stage organ failure, pathologic extracellular matrix accumulation, a fibrotic condition, disease states associated with immunosuppression (such as different forms of malignancy, chronic degenerative diseases, and AIDS), diabetic nephropathy, tumour growth, kidney damage (for example obstructive neuropathy, IgA nephropathy or non-inflammatory renal disease) or renal fibrosis.

34. A method of treating a patient in need of modulation of activin or TGF $\beta$  signalling, the method comprising administering to the patient an effective



amount of a polypeptide, molecule, compound, nucleic acid or antibody as defined in Claim 26.

35. A method of treating a patient with cancer the method comprising  
5 administering to the patient an effective amount of a polypeptide, molecule, compound, nucleic acid or antibody as defined in Claim 26.

36. A method of reducing extracellular matrix deposition or encouraging  
tissue repair and/or regeneration, or tissue remodelling or healing of a wound,  
10 injury or surgery, or reducing scar tissue formation arising from injury to the brain, the method comprising administering for the patient an effective amount of a polypeptide, molecule, compound, nucleic acid or antibody as defined in Claim 26.

15 37. A method of treating a disease or condition as defined in Claim 33, the method comprising administering to the patient an effective amount of a polypeptide, molecule, compound, nucleic acid or antibody as defined in Claim 26.

20 38. A substantially pure complex comprising (1) a Smad2 or Smad3 polypeptide, (2) a Smad4 polypeptide and (3) a Mixer and/or Milk and/or Bix polypeptide.

---

39. A preparation comprising (1) Smad2 or Smad3 polypeptide, (2) a Smad4 polypeptide and (3) a Mixer and/or Milk and/or Bix polypeptide (in the form of  
25 a complex or otherwise) when combined with other components *ex vivo*, said other components not being all of the components found in the cell in which said (1) Smad2 or Smad3 polypeptide, (2) a Smad4 polypeptide and (3) a Mixer

---

and/or Milk and/or Bix polypeptide (in the form of a complex or otherwise) are naturally found.

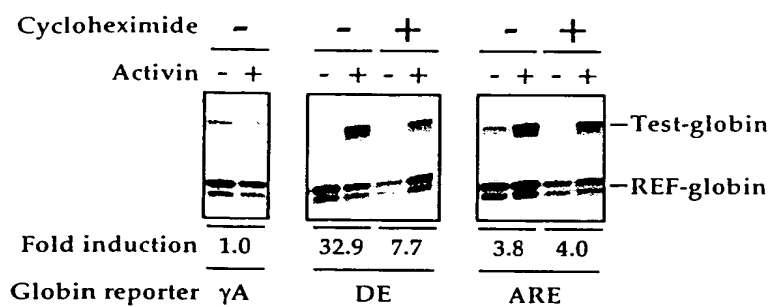
ABSTRACT

- A polypeptide (interacting polypeptide) capable of interacting with a Smad polypeptide wherein the interacting polypeptide comprises the amino acid sequence PP(T/N)K and is less than 150 amino acids in length or is not full-length *Xenopus* or human FAST1 or a fragment thereof, mouse FAST2, *Xenopus* Milk, *Xenopus* Mixer or *Xenopus* Bix2. The Smad polypeptide may be Smad2 or Smad3.
- 10 The interacting polypeptides are useful in screening assays and in medicine.

FIGURE 1



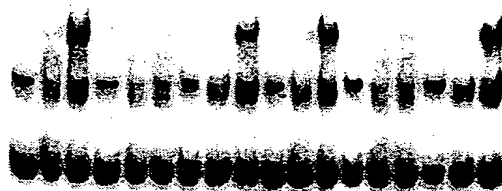
**A**



**B**

Competitor	-	DE wt		DE m1		DE m2		DE m3		ARE		
Activin injected	-	-	+	-	+	-	+	-	+	-	+	
Embryo stage	St8	St11	St8	St11	St8	St11	St8	St11	St8	St11	St8	St11

DEBP →



Probe

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18

DE wt CATTAAATCAGATTAAACGGTGAGCAATTAG

DE m1 --G-C-G--C-G-C-----

DE m2 --G-C-----G-C-----G-C-

DE m3 -----TTGT-----G-

**C**

Embryo stage	St 8	St 10.5
Cycloheximide	-	- + - +
Activin injected	-	- - + +

DEBP →



**Figure 1**



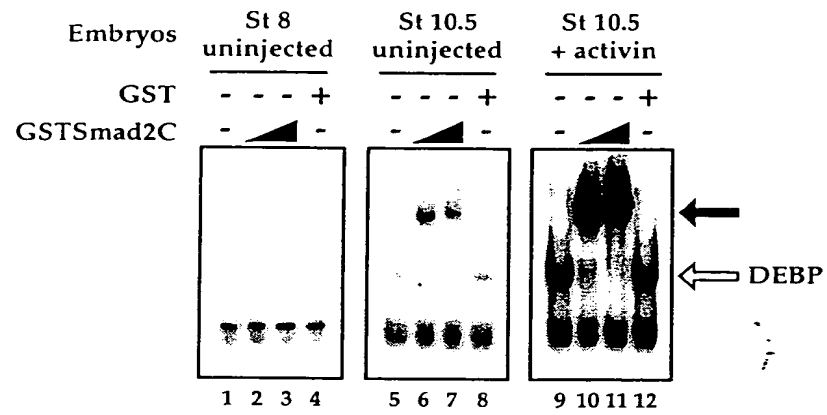
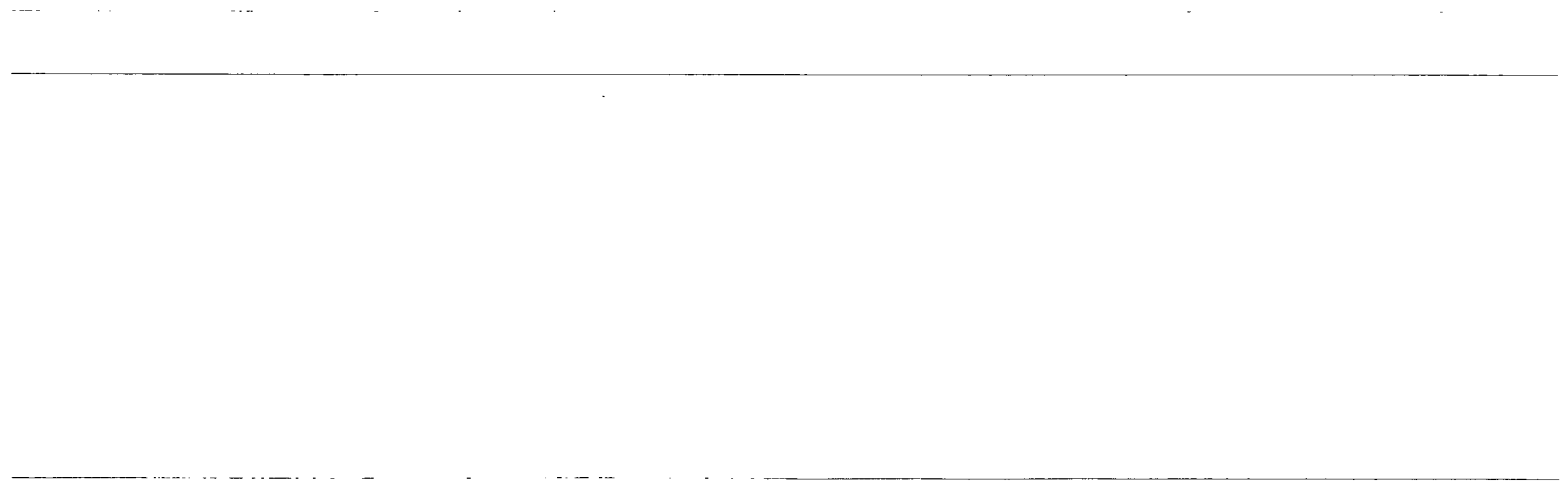


Figure 2





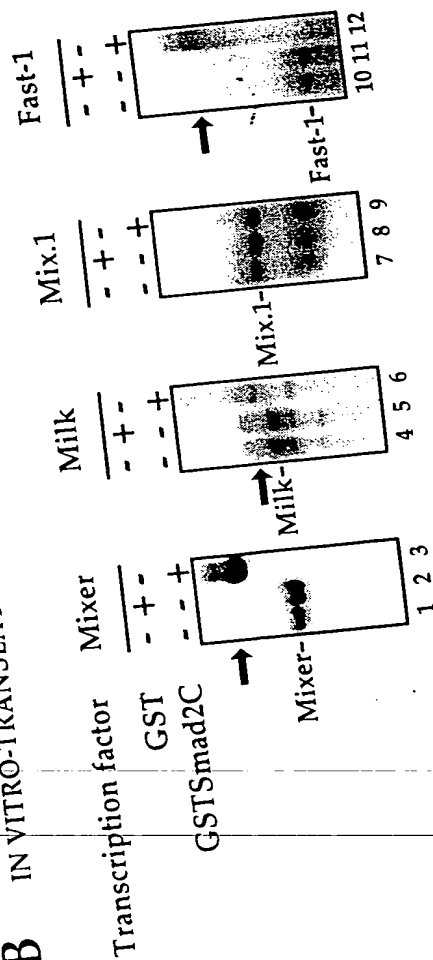
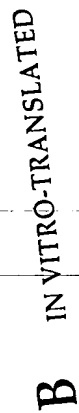
[illegible]

Figure 3



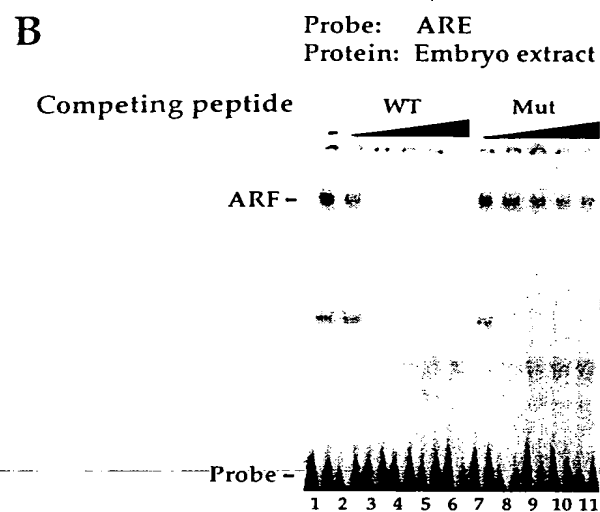
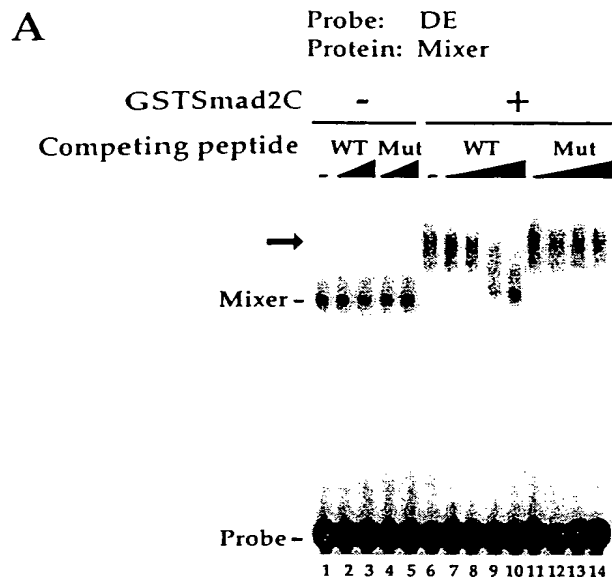
.

\_\_\_\_\_

\_\_\_\_\_

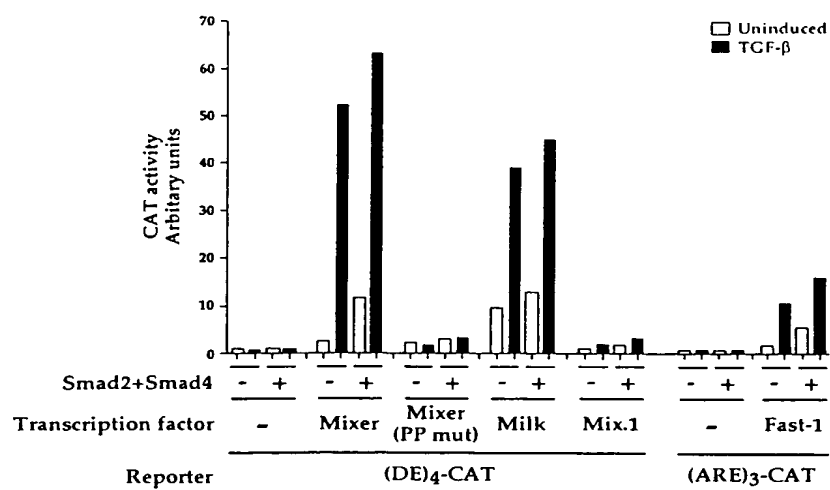
\_\_\_\_\_

\_\_\_\_\_





A



B

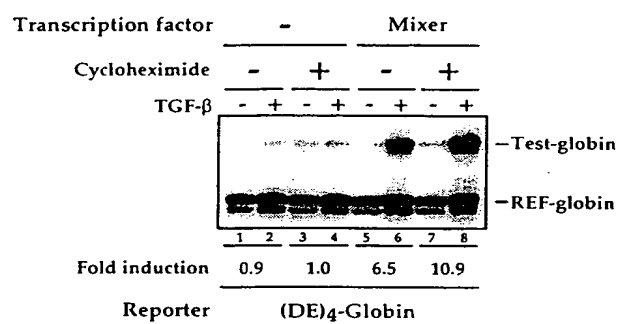
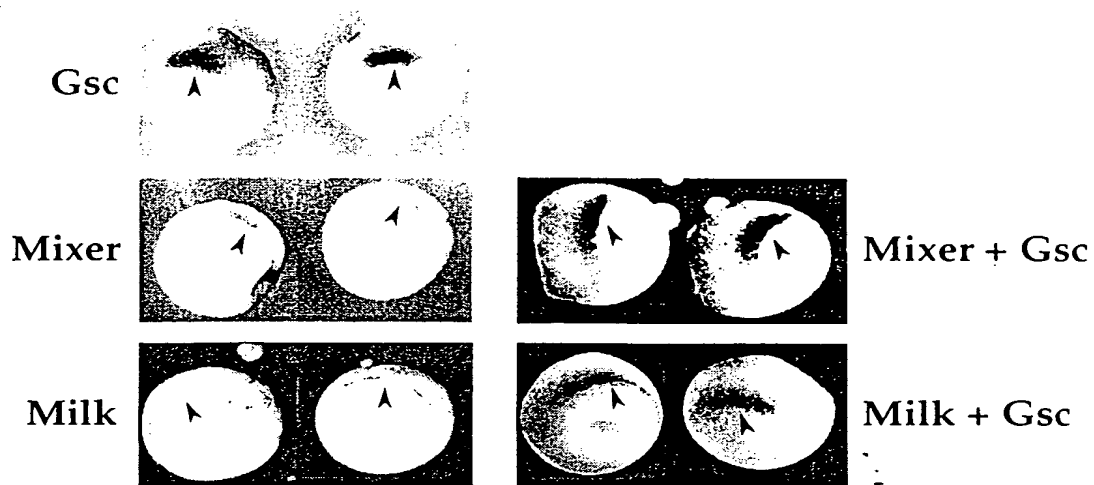


Figure 7



A



B

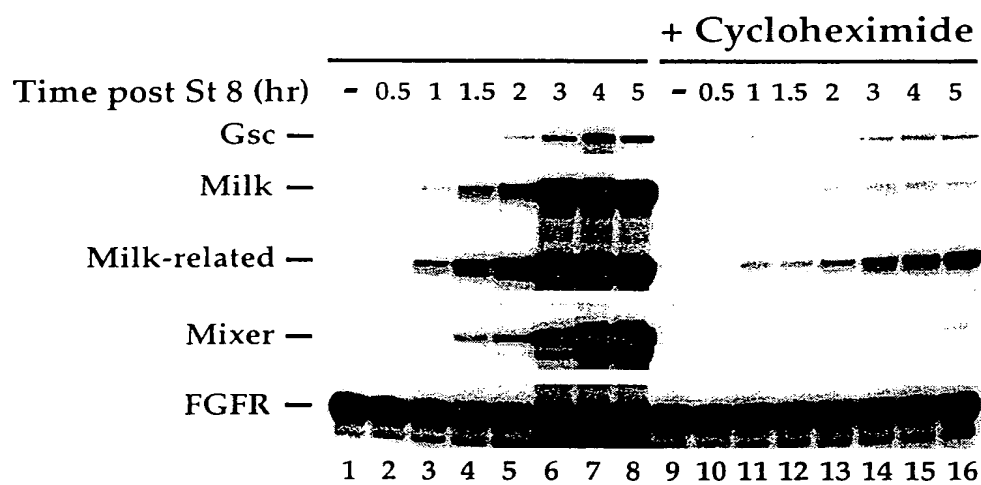


Figure 8A and B

---

---

---



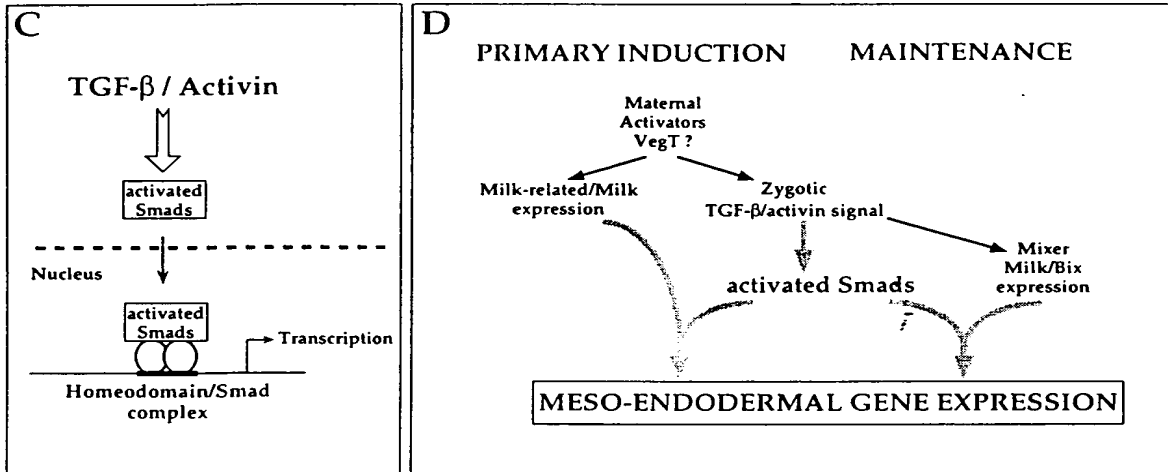


Figure 8 C and D



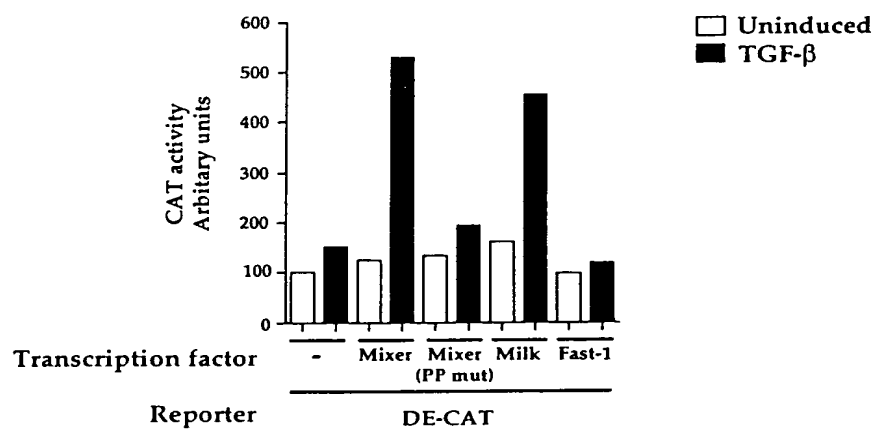
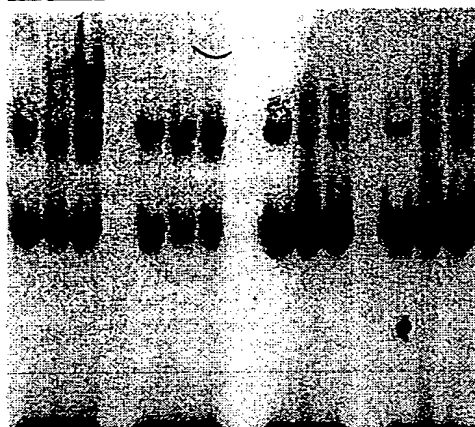


Figure 9



Probe	DE wt			DE m1			DE m2			DE m3		
Activin injected	-	-	+	-	-	+	-	-	+	-	-	+
Embryo stage	<u>St8</u>	<u>St11</u>		<u>St8</u>	<u>St11</u>		<u>St8</u>	<u>St11</u>		<u>St8</u>	<u>St11</u>	

DEBP ⇒



Probe

	$\xrightarrow{\hspace{1cm}}$ $\xleftarrow{\hspace{1cm}}$ $\xleftarrow{\hspace{1cm}}$
DE wt	CATTAATCAGATTAACGGTGAGCAATTAG
	-----
DE m1	--G-C-G---C-G-C-----
DE m2	--G-C-----G-C-----G-C-
DE m3	-----TTGT-----G---

Figure 10



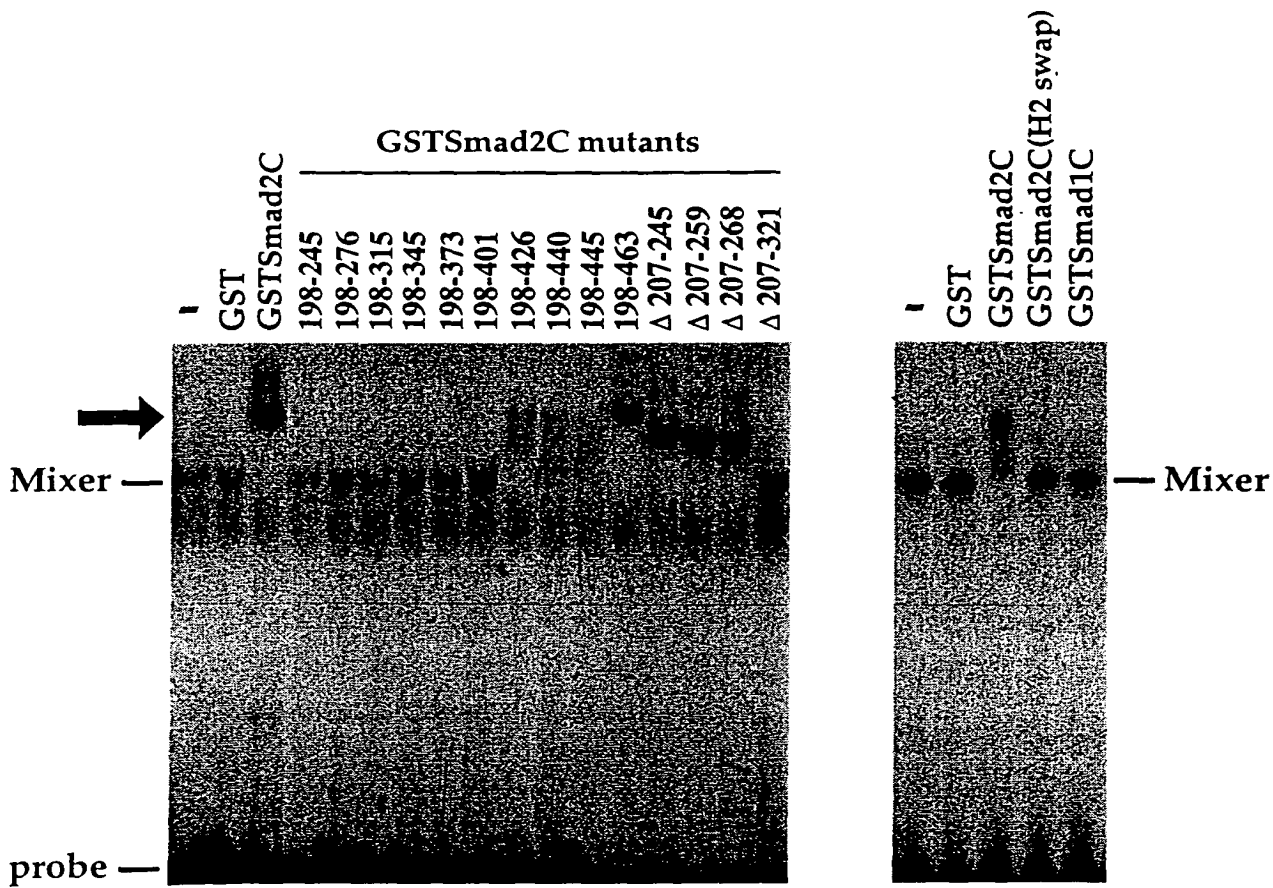


Figure 11





XSmad2

MAGSMSSILPFTPPVVKRLLGWKKKSASGTTGAGGDEQNGQEEKWCEKAVK  
SLVKKLKKKTGQLDELEKAITTQNCNTKCVTI PSTCSEIWGLSTANTIDQW  
DTTGLYSFSEQTRSLDGRLQVSHRKGLPHVIYCRLWRWPD L HSHHELKAI  
ENCEYAFNLKKDEV CVN PYHYQ RVETPVLPPVLVPRHTEILTEL PPLDDY  
THSIPENTNFPAGIEPQSNIYPETPPPGYI SEDGETSDQQLNQSM DTGSP  
AELSPSTLSPVNHNLDLQPV TYSEPAFWCSIAYYELNQRVGETFHASQPS  
LTVDGFTDPSNSERFCLGLLSNVNRNATVEMTRRHIGRGVRLYYIGGEVF  
AECLSDSAIFVQSPNCNQRYGWHPATVCKIPPGCNLKI FNNQEFAALLAQ  
SVNQGF EAVYQLTRMCTIRMSFVKGWGA EYRRQTVTSTPCWIELHLNGPL  
QWLDKVLTQMGSPSVRCSSMS

Hsmad2

MSSILPFTPPVVKRLLGWKKKSAGGSGGAGGGEQNGQEEKWCEKAVKSLVK  
KLKKTGR LDELEKAITTQNCNTKCVTI PSTCSEIWGLSTPNTIDQWDTTG  
LYSFSEQTRSLDGRLQVSHRKGLPHVIYCRLWRWPD L HSHHELKAIENCE  
YAFNLKKDEV CVN PYHYQ RVETPVLPPVLVPRHTEILTEL PPLDDYTHSI  
PENTNFPAGIEPQSNIYPETPPPGYI SEDGETSDQQLNQSM DTGSPAELS  
PTTLSPVNHSLDLQPV TYSEPAFWCSIAYYELNQRVGETFHASQPSLTV  
DGFTDPSNSERFCLGLLSNVNRNATVEMTRRHIGRGVRLYYIGGEVFAECL  
SDSAIFVQSPNCNQRYGWHPATVCKIPPGCNLKI FNNQEFAALLAQSVNQ  
GF EAVYQLTRMCTIRMSFVKGWGA EYRRQTVTSTPCWIELHLNGPLQWLD  
KVLTQMGSPSVRCSSMS

Xsmad3

MSSILPFTPPIVKRLLGWKKGEQNGQEEKWCEKAVKSLVKKLKKSGQLDE  
LEKALTTSISTKCITIPRSLDGRLQVSHRKGLPHVIYCRLWRWPD L HSH  
HEL RAMEVCEYAFSMKKDEV CVN PYHYQ RVETPVLPPVLVPRNTEIPAEF  
PSLDDYSHSIPENTNFPAGIEPQINIIYPETPPPGYLSE DGETSDQMNHSI  
DTGSPNLSPNSMSPAHSNMDLQPV TYCEPAFWCSISYYELNQRVGETFHA  
SQPSMTVDGFTDPSNSERFCLGLLSNVNRNA AVELTRRHIGRGVRLYYIG  
GEVFAECLSDNAIFVQSPNCNQRYGWHPATVCKIPPGCNLKI FNNQEFAA  
LLAQSVNQGF EAVYQLTRMCTIRMSFVKGWGA EYRRQTVTSTPCWIELHL  
NGPLQWLDKVLTQMGSPSIRCSSVS

Hsmad3

MSSILPFTPPIVKRLLGWKKGEQNGQEEKWCEKAVKSLVKKLKKKTGQLDE  
LEKAITTQNVNTKCITIPRSLDGRLQVSHRKGLPHVIYCPVRWPD L HSH  
EL RAMELCEFAFNMKKDEV CVN PYHYQ RVETPVLPPVLVPRHTEIPA EFP  
PLDDYSHSIPENTNFPAGIEPQSNIIYPETPPPGYLSE DGETSDHQMNHSM  
AGSPNLSPNPMSPAHNNDLQPV TYCEPAFWCSISYYELNQRVGETFHAS  
QPSMTVDGFTDPSNSERFCLGLLSNVNRNA AVELTRRHIGRGVRLYYIGG  
EVFAECLSDSAIFVQSPNCNQRYGWHPATVCKIPPGCNLKI FNNQEFAAL  
LAQSVNQGF EAVYQLTRMCTIRMSFVKGWGA EYRRQTVTSTPCWIELHLN  
GPLQWLDKVLTQMGSPSIRCSSVS

Figure 12 (page 1 of 2)



Full XSmad3 seq [301 to 1578] -> Translate • 1-frame

DNA sequence 1969 bp GAATTCGGCAGC ... CGTGCCGAATTC linear

```

301/1      331/11      361/21
ATG TCC TCG ATA CTT CCC TTC ACT CCC CCC ATT GTC AAG AGG CTG CTG GGA TGG AAG AAG GGA GAG CAG AAC GGA CAA GAG GAG AAA TGG
M S S I L P F T P P I V K R L L G W K K G E Q N G Q E E K W
391/31      421/41      451/51
TGC GAG AAG GCA GTG AAA AGT CTG GTG AAG AAG CTG AAG AGC AGC GGC CAG CTG GAC GAG CTG GAG AAG GCA CTG ACC ACC CAG AGC ATC
C E K A V K S L V K K L K K S G Q L D E L E K A L T T Q S I
481/61      511/71      541/81
AGC ACC AAG TGC ATC ACC ATC CCC AGG TCT CTG GAT GAG ACA CTC CAG GTG TCC CAT CGT AAG GGG TTG CCT CAT GTC ATA TAC TGC CGC
S T K C I T I P R S L D G R L Q V S H R K G L P H V I Y C R
571/91      601/101      631/111
CTC TGG AGG TGG CCA GAT CTG CAC AGT CAC CAC GAG CTG CGA GCC ATG GAG GTG TGT GAA TAT GCC TTC AGC ATG AAG AAG GAC GAA GTG
L W R W P D L H S H H E L R A M E V C E Y A F S M K K D E V
661/121      691/131      721/141
TGT GTG AAT CCT TAT CAC TAC CAG AGA GTG GAG ACT CCA GTT CTA CCT CCT GTG CTG GTT CCA CGA AAC ACC GAA ATT CCG GCA GAA TTC
C V N P Y H Y Q R V E T P V L P P V L V P R N T E I P A E F
751/151      781/161      811/171
CCT TCA CTT GAT GAT TAC AGC CAT TCC ATT CCG GAA AAT ACC AAT TTC CCA GCA GGC ATC GAG CCA CAG ATA AAC TAC ATT CCA GAA ACG
P S L D D Y S H S I P E N T N F P A G I E P Q I N Y I P E T
841/181      871/191      901/201
CCT CCT CCC GGT TAC TTG AGC GAA GAT GGG GAA ACA AGC GAC CAA ATG AAT CAC AGT ATA GAT ACA GGG TCT CCA AAT CTG TCA CCA AAC
P P P G Y L S E D G E T S D Q M N H S I D T G S P N L S P N
991/211      961/221      991/231
ATG TCT CCT GCT CAT AGC AAC ATG GAC CTG CAG CCT GTC ACA TAC TGC GAG CCG GCC TTT TGG TGT TCC ATC TCC TAC TAT GAG CTT
S M S P A H S N M D L Q P V T Y C E P A F W C S I S Y Y E L
1021/241      1051/251      1081/261
AAC CAA CGC GTA GGG GAG ACG TTC CAC GCT TCC CAG CCC TCC ATG ACA GTG GAT GGA TTC ACC GAT CCT TCC AAC TCT GAA CGT TTC TGC
N Q R V G E T F H A S Q P S M T V D G F T D P S N S E R F C
1111/271      1141/281      1171/291
CTG GGG CTG TTG TCC AAC GTA AAT CGG AAT GCA GCT GTG GAG CTG ACA CGG AGA CAC ATC GGG AGA GGC GTG CGG CTG TAT TAC ATT GGA
L G L L S N V N R N A A V E L T R R H I G R G V R L Y Y I G
1201/301      1231/311      1261/321
GGG GAA GTG TTT GCC GAG TGC CTC AGT GAC AAT GCC ATA TTT GTA CAG TCC CCA AAT TGT AAC CAG CGC TAC GGT TGG CAT CCT GCC ACA
G E V F A E C L S D N A I F V Q S P N C N Q R Y G W H P A T
1291/331      1321/341      1351/351
GTC TGC AAG ATT CCA CCA GGC TGT AAC CTG AAG ATA TTT AAT AAC CAG GAG TTT GCT GCT CTT TTG GCT CAG TCA GTA AAC CAG GGC TTT
V C K I P P G C N L K I F N N Q E F A A L L A Q S V N Q G F
1381/361      1411/371      1441/381
GAG GCT GTG TAT CAG CTT ACG AGG ATG TGC ACC ATA CGC ATG AGT TTC GTC AAA GGC TGG GGA GCC GAA TAC AGG CGA CAG ACT GTG ACT
E A V Y Q L T R M C T I R M S F V K G W G A E Y R R Q T V T
1471/391      1501/401      1531/411
AGC ACC CCC TGC TGG ATC GAG CTG CAC TTG AAC GGG CCC TTG CAA TGG TTG GAT AAG GTT CTC ACT CAG ATG GGG TCT CCA AGT ATC CGC
S T P C W I E L H L N G P L Q W L D K V L T Q M G S P S I R
1561/421
TGC TCC AGT GTT TCT TAA
C S S V S *

```

Figure 12 (page 2 of 2)



Bix.1 ORF [1 to 1206] -> Translate • 1-frame

DNA sequence 1206 bp atgcttgatgat ... actctattctga linear

1/1	31/11	61/21
atg ctt gga tat acc caa ggg atg gag cag ctc tat agc acc tac ttc tcc tcc agt gac	ccc tca atg ggc ttc agt tct gcc ctg gac	
M L G Y T Q G M E Q L Y S T Y F S S S D P S M G F S S A L D		
91/31	121/41	151/51
tcc ctc atg gga atg ggt ttc cct cct gta caa cag agt cca gta cag caa gct aat gtg	aaa gga gac atg aag gct gga gat ctt gga	
S L M G M G F P P V Q Q S P V Q Q A N V K G D M K A G D L G		
181/61	211/71	241/81
gcc aat caa aca aca cag cac aag gaa gct acc aat cag cag aag gtt tcc ccc aca cag	atg tcc aat cgc agg aag aga act gtc tac	
A N Q T T Q H K E A T N Q Q K V S P T Q M S N R R K R T V Y		
271/91	301/101	331/111
agt ccc tca gat ctg gcc cga ctg gag cag tac ttc aga act aat atg tac cca gac atc	cac cag cgg gaa gaa atg gcc aga caa atc	
S P S D L A R L E Q Y F R T N M Y P D I H Q R E E M A R Q M		
361/121	391/131	421/141
ggc ttg ccc gag tca cgc ata cag gtt tgg ttc caa aac agg agg tca aaa gcc aga cgc	caa gga tcc aga tct acc aag ttg gct gct	
G L P E S R I Q V W F Q N R R S K A R R Q G S R S T K L A A		
451/151	481/161	511/171
atg gga gat tac tat aac agc acc cca aag tac aat gca gca cct gtc tcc aat ggc acc	ata aat gta cca cag caa cag cgg atg tta	
M G D Y Y N S T P K Y N A A P V S N G T I N V P Q Q Q R M L		
541/181	571/191	601/201
tcc tac cag cat cat gcc cag cca ttg gac act cta cac tat ggc ttc cac cca aat gtc	tcc atg caa gga acc agt cag agc agg atc	
S Y Q H H A Q P L D T L H Y G F H P N V S M Q G T S Q S R M		
631/211	661/221	691/231
tca tca gca cca gct ccc cat cat tca ggg aat gtg tct cag aag cat cca atg ggt	ttt cca cag caa taa gtc caa ccc att ctg	
Y S S A P A P H H S G N V S Q K H P M G F P Q Q Q V Q P I L		
721/241	751/251	781/261
gat tta cag cag aat tac ttc ccc ttt tct gag gac ccc ctg tct tgc cct gaa tcc tca	tgg gct gcc cca aac cag aga ctg cca gtt	
D L Q Q N Y F P F S E D P L S C P E S S W A A P N Q R L P V		
811/271	841/281	871/291
cat ctg aca gca tct agt acc tac cct cac aat gac ttg cct agt aac ccc act gca agc	cac tcc acc aag gga atg agc cct ggc aag	
H L T A S S T Y P H N D L P S N P T A S H S T K G M S P G K		
901/301	931/311	961/321
gag act gag aga gac ttt ggt ggg aga cat aac caa gta tcc atg cat tct aac ctt atg	atg gac att agc aac ttc cca cct acc aag	
E T E R D F G G R H N Q V S M H S N L M M D I S N F P P T K		
991/331	1021/341	1051/351
acc atc acc gcc aat atg aac acc atc atc cca cag atg cca ggg gct tca tgc tgg agc	agc cat gag gat atc aat gcc tac tct aca	
T I T A N M N T I I P Q M P G A S C W S S H E D I N A Y S T		
1081/361	1111/371	1141/381
cag ggg gca gtg ccc atg gca ggt tgt agc ccc tat gga cat ggt tct cct gat tca gga	gtc agt gac aca tct aca gag tca gtc tca	
Q G A V P M A G C S P Y G H G S P D S G V S D T S T E S V S		
1171/391	1201/401	
gac tgg gaa gag aat att ata aga act cta ttc tga		
D W E E N I I R T L F *		

Figure 13 (page 18 of 12)



Bix2 ORF [1 to 1206] -> Translate • 1-frame

DNA sequence 1206 bp atgcttggatg ... actctattctga linear

```

1/1
atg ctt gga tat acc caa ggg atg gag cag ctc tac agc acc tac ttc tcc tcc agt gac cat cta atg agc tcc agg tct gcc ctg gac
M L G Y T Q G M E Q L Y S T Y F S S S D H L M S S R S A L D
91/31
tcc ctc atg gga atg ggt ttc cct cct gta caa cag aat cca gta cag caa gtt aat gtt aaa gca gac atg aag gct ggg gat ctt aga
S L M G M G F P P V Q Q N P V Q Q V N V K A D M K A G D L R
181/61
gcc aat cag aca gaa cag cac aag gaa gcc tcc aat cag cag aaa gtt tcc ccc aca ttg atg tcc aat cgc agg agg aga act gtc tac
A N Q T E Q H K E A S N Q Q K V S P T L M S N R R R R T V Y
271/91
agt ccc tca gat ctg gcc cga ctg gag cag tac ttc caa aca aat atg tac cca gac atc cac caa cgg gag gaa ctg gcc aga caa atg
S P S D L A R L E Q Y F Q T N M Y P D I H Q R E E L A R Q M
361/121
ggc ttg ccc gag tct cgc att cag gtt tgg ttc cag aac agg agg tca aaa gcc aga cgc caa gga tcc aga tct acc aag ccg ggg tct
G L P E S R I Q V W F Q N R R S K A R R Q G S R S T K P G S
451/151
gtg gaa tat tac tat aac agc acc ccg atg tac ttc tca gca ccc aca gcc aat ggc aca atc acc ata gca cag caa cag cag atg ttg
V E Y Y Y N S T P M Y F S A P T A N G T I T I A Q Q Q Q M L
541/181
ttc tac cag cag cag cag gtc cag cca ttg acc act cta cac aat ggc ttc cac cca aaa gtc tcc atg caa gga acc aat cag aac aag
Y Q Q Q V Q P L T T L H N G F H P K V S M Q G T N Q N K
61/21
cat cta atg agc tcc agg tct gcc ctg gac
H L M S S R S A L D
151/51
aaa gca gac atg aag gct ggg gat ctt aga
K A D M K A G D L R
241/81
atg tcc aat cgc agg agg aga act gtc tac
M S N R R R R T V Y
331/111
cac caa cgg gag gaa ctg gcc aga caa atg
H Q R E E L A R Q M
421/141
caa gga tcc aga tct acc aag ccg ggg tct
Q G S R S T K P G S
511/171
atc acc ata gca cag caa cag cag atg ttg
I T I A Q Q Q Q M L
601/201
gtc tcc atg caa gga acc aat cag aac aag
V S M Q G T N Q N K
691/231
ttt cca cag cat caa gtc caa ccc att atg
F P Q H Q V Q P I M
781/261
ggg gca gcc ccc aac cag aga ctc cca ttg
G A A P N Q R L P L
871/291
cac tgc acc cag ggt atg atc cct aac aag
H C T Q G M I P N K
961/321
atg gac ttt cca cct aac aag acc atc acc
M D F P P N K T I T
1051/351
gat gcc tac tct aca aca agg
D A Y S T T R
1141/381
gca tat gca gag tca gtg tca
A Y A E S V S
1201/401
gac tgg gaa gag aat att atc aag act cta ttc tga
D W E E N I I K T L F

```

Bix2

This is almost identical to Milk.

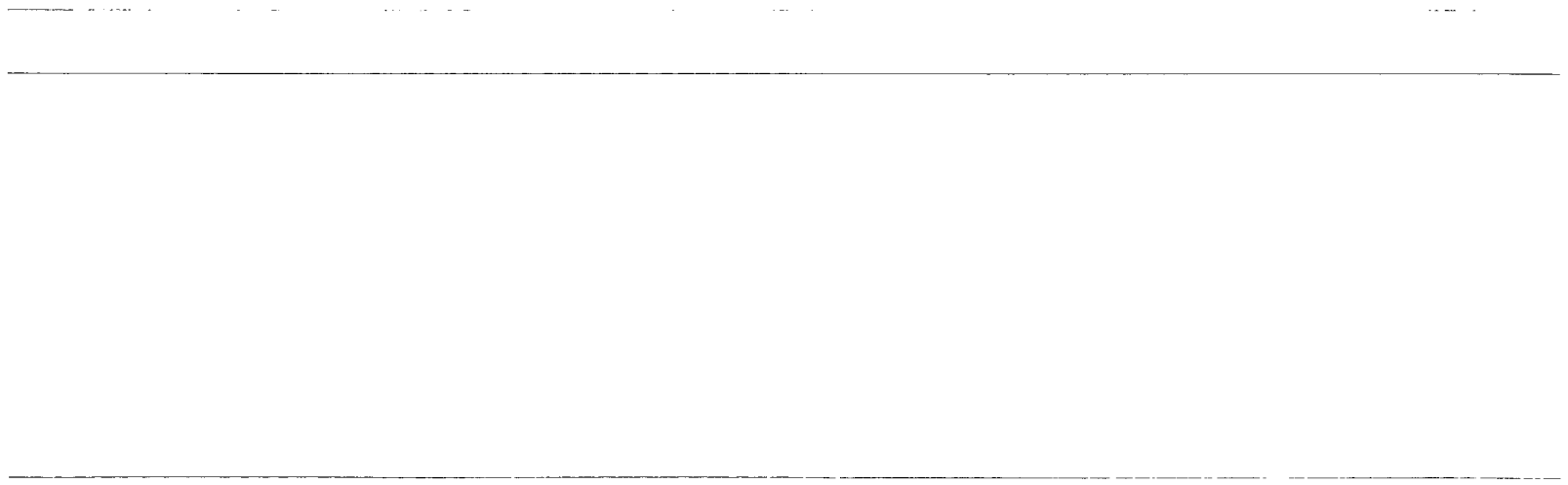
26 aa region encompassing PTHu marker.

316-341.

Figure 13 (page 2 of 12)



1





Bix.3 ORF [1 to 1170] -> Translate • 1-frame

DNA sequence 1170 bp ATGCTTGGATAT ... TCTCTATTTTGA linear

```

1/1          31/11          61/21
ATG CTT GGA TAT ACC CAA GAG ATG gag cag ctc tat agc acc tac ttc tcc tcc agt gac ccc tca atg ggc ttc agt tct gcc ctg gac
M L G Y T Q E M E Q L Y S T Y F S S S D P S M G F S S A L D
91/31          121/41          151/51
tcc ctc atg gga atg ggt ttc cct cct gta cca cag aat cca gta cag caa gct aat gtg aaa gga cac atg aag gct gga gat ctt gga
S L M G M G F P P V Q Q G N P V Q Q A N V K G H M K A G D L G
181/61          211/71          241/81
gcc aat caa aca aca cag cac aag gaa gct tcc aat cag cag aag gtt tcc ccg aca cag atg tcc aat cgc agg aag aga act gtc tac
A N Q T T Q H K E A S N Q Q K V S P T Q M S N R R K R T V Y
271/91          301/101          331/111
agt ccc tca gat ctg gcc cga ctg gag cag tac ttc aga aca aat atg tac cca gat atc cac cag cgg gaa gaa atg gcc aga caa atg
S P S D L A R L E Q Y F R T N M Y P D I H Q R E E M A R Q M
361/121          391/131          421/141
ggc ttg ctc gag tca cgc ata cag gtt tgg ttt cag aac agg agg tca aaa gcc aag cgt caa ggg tcc aga tct acc aag ttg gct gct
G L L E S R I Q V W F Q N R R S K A K R Q G S R S T K L A A
451/151          481/161          511/171
gtt gga gat tac tat aac aga acc cca atg tac aac cca gca ccc aca gcc aat ggc aca ata act gta gca cag caa caa cga gtg ata
V G D Y Y N R T P M Y N P A P T A N G T I T V A Q Q Q R V I
541/181          571/191          601/201
tct tat cag cag cag gtc cag cca ttg gcc act cta cac tat ggc ttc cag cca aat gtc tcc atg caa gga acc agt ctg gac aag atg
S Y Q Q Q V Q P L A T L H Y G F Q P N V S M Q G T S L D K M
61/211          661/221          691/231
tat tca tct cag cag aat cca atc agc ttt cca cag caa caa gtc caa ccc att atc aat gta cag cag aat tac ttc cac cag ccc tgg
Y S S Q Q N P I S F P Q Q Q V Q P I I N V Q Q N Y F H Q P W
721/241          751/251          781/261
gac ctc ctg cct tgc cct gaa tcc tca tgg aca gtc aac agc cag aga ctc cca gga cat cca aca aca tcc agt acc cac cct cat att
D L L P C P E S S W T V N S Q R L P G H P T T S S T H P H I
811/271          841/281          871/291
gat ttg cct agt aag ccc att aca agc cac tgc acc aag ggt atg agt cct ggc aag gaa tct gag aga gac ttt ggt tgg aga caa aac
D L P S K P I T S H C T K G M S P G K E S E R D F G W R Q N
901/301          931/311          961/321
caa gta acc atg cat tct aac ctc gtg atg gaa ttc cca cct aac aag acc ata acc cct gat atg aac acc atc atc cca cag ata cca
Q V T M H S N L V M E F P P N K T I T P D M N T I I P Q I P
991/331          1021/341          1051/351
ggg gca aca ggt tgg aag aac cag gag gat atc aat acc tac tct aca cag ggg gca ctg tcc agg gca ggg tgt agc agc tat gga ctt
G A T G W K N Q E D I N T Y S T Q G A L S R A G C S S Y G L
1081/361          1111/371          1141/381
cac tct ccg tcc tca gac tca gga gtc agt gat gca tct aca gag tca gtc tca gac tgg gaa gag aac CTT CTG AAA TCT CTA TTT TGA
H S P S S D S G V S D A S T E S V S D W E E N L L K S L F *

```

Bix.3

PPNII motif v. similar to Milk/Mixie.

305-319.

Figure 13 (page 3 of 12)



Bix4 ORF [1 to 1164] -> Translate \* 1-frame

DNA sequence 1164 bp ATGCTTGGATAT ... TTCACCCCTGGA linear

```

1/1                               31/11                               61/21
ATG CTT GGA TAT ACC CAA GGG ATG gag cag ctc tac agc acc tac ttc tcc tcc agt gac cac ctc atg agc tcc agg tct gcc ctg gac
M L G Y T Q G M E Q L Y S T Y F S S S D H L M S S R S A L D
91/31                               121/41                               151/51
tcc ctc gtg ggt tgc ggt ttc cct cct gta caa cag aat cca gta cag caa gtt aat gta aaa gac atg aag gct ggg gaa cnc gac aag
S L V G S G F P P V Q Q N P V Q Q V N V K D M K A G E H D K
181/61                               211/71                               241/81
gaa gcc gcc aat cag cag aaa gtt tcc ccc aca ctg atg tcc aat cgc agg agg aga act gtc tac agt ccc tca gat ctg gcc aga ctg
E A A N Q Q K V S P T L M S N R R R R R T V Y S P S D L A R L
271/91                               301/101                               331/111
gag cag aac ttc caa act aat atg tac cca gac atc cac cag cgg gag gaa atg gcc agg caa atg ggc ttg ccc gag tct cga gtt cag
E Q N F Q T N M Y P D I H Q R E E M A R Q M G L P E S R V Q
361/121                               391/131                               421/141
gtt tgg ttc cag aac agg aga tca aaa gcc aga cgc caa gga tcc aga tcc acc aag ccg gct ggt gtg gga gat tac tat aac agc acc
V W F Q N R R S K A R R Q G S R S T K P A G V G D Y Y N S T
451/151                               481/161                               511/171
cca atg tac aac cca gca ccc aca gcc aat tgc aca agc cct gta gca cag cag cgt atg ttg tcc tac cag cag cca ttg gcc act cta
P M Y N P A P T A N C T S P V A Q Q R M L S Y Q Q P L A T L
541/181                               571/191                               601/201
cac tat ggc ttc cac cca aat gtc acc atg caa gga acc aat cag agc aag tat tca tca gca tca gct cca cat cca ggg aat gtg tct
H Y G F H P N V T M Q G T N Q S K Y S S A S A P H P G N V S
631/211                               661/221                               691/231
ctg cac cgg atg ggt ttt caa cag cca gtc caa ccc att cag cag aat tat ttc cac ttg tcc cag gac ctc ctg tct tgc cct gaa
L H R M G F Q Q P V Q P I Q Q N Y F H L S Q D L L S C P E
721/241                               751/251                               781/261
tcc tca tgg gca gcc ccc aac caa agg cgc cca gta cat ccg aca gca tct agt acc tac ccg cac agt tac caa cct agt aag ccc ctt
S S W A A P N Q R R P V H P T A S S T Y P H S Y Q P S K P L
811/271                               841/281                               871/291
aca ggc cac tat acc cag ggt atg agc cct ggc ttt gag act gag aga gac ctc ggt gga aga cat aac caa gtt tcc atg cat tct aac
T G H Y T Q G M S P G F E T E R D L G G R H N Q V S M H S N
901/301                               931/311                               961/321
atc atg atg gat ttt agc aac ttc caa ccc aag aag acc atc acc ccc gat atg aac acc atc atc cca cag atc cca gat gct aca ggc
I M M D F S N F Q P K K T I T P D M N T I I P Q I P D A T G
991/331                               1021/341                               1051/351
tgg agc aac cag gag ggt act gat gcc tac tct aca cag ggg gca ctg ccc agg gca caa tgt agc ccc tat gga cat gga tat cct gcc
W S N Q E G T D A Y S T Q G A L P R A Q C S P Y G H G Y P A
1081/361                               1111/371                               1141/381
tca gac tca gga gtg agt gac aca tct aca gag tca atc tca gac tgg gaa gag aat att ata AGA ACT CTA TTC TGC GCC TGA
S D S G V S D T S T E S I S D W E E N I I R T L F C A *

```

Figure 13 (page 4 of 12)



Mixer ORF Strider [1 to 1116] -> Translate • 1-frame

DNA sequence 1116 bp atggacacgttc ... gacaacctgtaa linear

```

1/1          31/11          61/21
atg gac acg ttc agc caa caa ctg gag gac ttc tac cca tct tgc ttc tct gcc agg tcc agc cca gtg ggc ttc act gat cca cca gcc
M D T F S Q Q L E D F Y P S C F S A R S S P V G F T D P P A
91/31          121/41          151/51
cag cac tta acc atg agc ctt ggt gcc att cag aag gat ttc caa gag tcc agc ttg aag gca aac gtc cag cca gtc agt gat cct caa
Q H L T M S L G A I Q K D F Q E S S L K A N V Q P V S D P Q
181/61          211/71          241/81
aca cta ggg acc cag aag tcc acc cct cct aca aaa caa gaa atg gtc tcc ccg gta tct tca gtt gat gtg aca ttg gga tca caa cgc
T L G T Q K S T P P T K Q E M V S P V S S V D V T L G S Q R
271/91          301/101          331/111
cgc aag aga aca ttt tac agc cag aac aag ctg gat gtt cta gaa cag ttc ttc cag acc aac atg tat cca gat att cac cac cgg gaa
R K R T F Y S Q N K L D V L E Q F F Q T N M Y P D I H H R E
361/121          391/131          421/141
gaa ctg gct aaa cgc att tac atc cca gag tcc aga gtt cag gtc tgg ttc cag aac aga aga gca aag gag cgc cgc gat aaa gcc aaa
E L A K R I Y I P E S R V Q V W F Q N R R A K E R R D K A K
451/151          481/161          511/171
tta aac ccc tcg cca gca gta ggc gtg tgc tac ccc agt ctt cga caa ccc aat aaa gaa atg tac ccc tcc aac aac cca aca cca aat
L N P S P A V G V C Y P S L R Q P N K E M Y P S N N P T P N
541/181          571/191          601/201
gtg cct gtt tcc cag caa cac atg gtt ttc caa aaa cca cag ggt caa ctg ttc atg aat tcc cag cag aat ccg ttc cag cca act cag
V P V S Q Q H M V F Q K P Q G Q L F M N S Q Q N P F Q P T Q
/211          661/221          691/231
tct cag ctt tgt tca gaa tcc acc tac gct gtt tcc cag cag agg atc ctg atg cag cag gct gca ccg agc agc tac cat gga att
E S Q L C S E S T Y A V S Q Q R I L M Q Q A A R S S Y H G I
721/241          751/251          781/261
tca gca tct tat aaa cct aca gac act cag cag cac ttc tac tca tac atg agc cca atg gga ggc acc cag gag aaa gtc atg gat cta
S A S Y K P T D T Q Q H F Y S Y M S P M G G T Q E K V M D L
811/271          841/281          871/291
agc aag aag cac agt cag atg ccc ttc cat ccc agt ctt cta atg gac ttc aac aac ttc cct ccc aac aag acc atc act cca gat atg
S K K H S Q M P F B P S L L M D F N N F P P N K T I T P D M
901/301          931/311          961/321
aat gtt aga atc cca cca att cct gtc tct gca ccg tca aac aac cac agt cgg atg aat gtc ttt aat acc aaa gag gcc ggc cca ttg
N V R I P P I P V S A P S N N H S R M N V F N T K E A G P L
991/331          1021/341          1051/351
gtg tcc ttg cca gag gat gtc tat gag gaa ttc tct ccg gtc tct gat tct ggt gtt agt gat gga tct acc atg tct ttg aca gac ttt
V S L P E D V Y E E F S P V S D S G V S D G S T M S L T D F
1081/361          1111/371
aaa gat aat gat gga tct gtg ctt gac aac ctg taa
K D N D G S V L D N L

```

This is no defined region that works

Mix 283-307

Figure 13 (page 5 of 12)



##NA Strider<sup>TM</sup> 1.3f2 ### Tuesday, August 3, 1999 12:26:40 pm

Milk ORF [1 to 1203] -> Translate • 1-frame

DNA sequence 1203 bp ATGCTTGGATAT ... ACTCTATTCTGA linear

```
1/1
ATG CTT GGA TAT ACC CAA GGG ATG GAG CAG CTC TAC AGC ACC TAC TTC TCC TCC AGT GAC CAT CTA ATG AGC TCC AGG TCT GCC CTG GAC
M L G Y T Q G M E Q L Y S T Y F S S S D H L M S S R S A L D
91/31
TCC CTC ATG GGA ATG GGT TTC CCT CCT GTA 121/41 CAA CAG AAT CCA GTA CAG CAA GTT AAT GTT 151/51 AAA GCA GAC ATG AAG GCT GGG GAT CTT AGA
S L M G M G F P P V Q Q N P V Q Q V N V K A D M K A G D L R
181/61
GCC AAT CAG ACA GAA CAG CAC AAG GAA GCC 211/71 TCC AAT CAG CAG AAA GTT TCC CCC ACA TTG 241/81 ATG TCC AAT CGC AGG AAG AGA ACT GTC TAC
A N Q T E Q H K E A S N Q Q K V S P T L M S N R R K R T V Y
271/91
AGT CCC TCA GAT CTG GCC CTT CTG GAG CAG TAC TTC CAA ACA AAT ATG TAC CCA GAC ATC CAC CAA CGG GAG GAA CTG GCC AGA CAA ATG
S P S D L A L L E Q Y F Q T N M Y P D I H Q R E E L A R Q M
361/121
GGC TTG CCC GAG TCT CGC ATT CAG GTT TGG TTC CAG AAC AGG AGG TCA AAA GCC AGA CGC CAA GGA TCC AGA TCT ACC AAG CCG GGG TCT
G L P E S R I Q V W F Q N R R S K A R R Q G S R S T K P G S
451/151
GTG GAA TAT TAC TAT AAC AGC ACC CCG ATG TAC AAC TCA GCA CCC ACA GCC AAT GGC ACA ATC ACC ATA GCA CAG CAA CAG CAG ATG TTG
V E Y Y Y N S T P M Y N S A P T A N G T I T I A Q Q Q Q Q M L
541/181
TCC TAC CAG CAG CAG CAG GTC CAG CCA TTG ACC ACT CTA CAC AAT GGC TTC CAC CCA AAA GTC TCC ATG CAA GGA ACC AAT CAG AAC AAG
Y Q Q Q Q V Q P L T T L H N G F H P K V S M Q G T N Q N K
61/21
ATT TAT TCA TCG GCA CCA GCT CCA AAT TCA GGG AAG GTG TTT CAG CAG CAT TCA GTG GGG TTT CCA CAG CAT_CAA GTC CAA CCC ATT ATG
I Y S S A P A P N S G K V F Q H S V G F P Q H Q V Q P I M
721/241
AAT TTA CAG CAG AAT TAC TAC CAC CAG CCC CAG GAC CTC CTG TCC TCA TGG TCC CCC TCA TGG GCA GCC CCC AAC CAG AGA CTC CCA TTG
N L Q Q N Y Y H Q P Q D L L S S W S P S W A A P N Q R L P L
811/271
CAT CTA ACA GCA TCT AGT ATC TAC CCT CAC AGT TAC CTG CCT AGT GAG CTC ATT ACA AGC CAC TGC ACC CAG GGT ATG ATC CCT AAC AAG
H L T A S S I Y P H S Y L P S E L I T S H C T Q G M I P N K
901/301
GAG ATT GAG GGA GAA CTA TGT GGA AGA CAT AAC CAA GTA TCC ATG CAT TCC AAC CTT ATG ATG GAC TTT CCA CCT AAC AAG ACC ATC ACC
E I E G E L C G R H N Q V S M H S N L M M D F P P N K T I T
991/331
CCC GAT ATG AAC ACC ATC ATC CCA CAG ATC ACA GAT GCT ACA GGC TGG AGC AGC CAG GAG GGT ACC GAT GCC TAC TCT ACA CAG GGG GCA
P D M N T I I P Q I T D A T G W S S Q E G T D A Y S T Q G A
1081/361
CTG CCC AGG GCA CAA TGT AGC CCC TAT GGA CAG GGA TCT CCT GCC TCA GAC TCA GGA ATC AGT GAC GCA TAT GCA GAG TCA GTG TCA GAC
L P R A Q C S P Y G Q G S P A S D S G I S D A Y A E S V S D
1171/391
TGG GAA GAG AAT ATT ATC AAG ACT CTA TTC TGA
W E E N I I K T L F *
```

Milk 316-340

equine 25da Reg Mixes.

Figure 13 (page 6 of 12)





Mix.1 ORF [69 to 1202] -> Translate • 1-frame

DNA sequence 1202 bp caactgagcagc ... ctcaaccttttaa linear

```

69/1          99/11          129/21
atg gat gga ttc agc caa caa ctg gag gac ctc tac cct tct tgc ttt tcc ccc tgc ccc agc cct ttg ggg ttc agt gag cca gtg att
M D G F S Q Q L E D L Y P S C F S P C P S P L G F S E P V I
159/31       189/41       219/51
cag cca ttt gcc atg aat ctg gcg cct gca gca cag aag gac ttc cag cag cat ccc aat agg aga gaa gtc acc aag atc cct gga gcc
Q P F A M N L A P A A Q K D F Q Q H P N R R E V T K I P G A
249/61       279/71       309/81
ggg gag cag agc cca gtc caa aat gtc aga ccc aaa gac gca att aat ccc aaa gag gct gac cca aga agc cca acg gct gat gcc tct
G E Q S P V Q N V R P K D A I N P K E A D P R S P T A D A S
339/91       369/101      399/111
ttg gtc cca gca tct cag cgc agg aaa agg acc ttc ttc acc cag gca caa ctg gat atc ctc gag cag ttc ttc caa acc aac atg tac
L V P A S Q R R K R T F F T Q A Q L D I L E Q F F Q T N M Y
429/121      459/131      489/141
cca gac atc cac cac cga gag gag cta gcg agg cac atc tat atc cca gag tcc cgc att cag gtc tgg ttc cag aac aga aga gca aag
P D I H H R E E L A R H I Y I P E S R I Q V W F Q N R R A K
519/151      549/161      579/171
gtc aga cgt caa ggt gca aaa gcc acc aag ccc att ctt gca ggc cat cat tat tcc ggc acc tct ggg gct aac agg gca atg ttt cct
V R R Q G A K A T K P I L A G H H Y S G T S G A N R A M F P
609/181      639/191      669/201
tca gca cct gcc cct aat agt tcc tca cat cag atg act aca tcc cgg gca cag gtc cag cct ctg aaa gaa tcc caa atg aat atg ttt
A P A P N S S S H Q M T T S R A Q V Q P L K E S Q M N M F
699/211      729/221      759/231
cac caa aac cag gga ttc ctt ecc tac cca gat tca tct tgt aat gtc tca agg cag agg ttc ctc atg tcc cag cca aca cct ggt gcc
H Q N Q G F L P Y P D S S C N V S R Q R F L M S /Q P T P G A
789/241      819/251      849/261
tac cac ctg ccc cag gca tca tcc aat gtc tac aat cag aat gtg aaa tct caa aac cct ctg tgg aac caa cag aca acc aac aga gtt
Y H L P Q A S S N V Y N Q N V K S Q N P L W N Q Q T T N R V
879/271      909/281      939/291
tat aca aat atg atg gag acg gtg ttg gac ctc agc aga aga ccc tcc cag cag atg cca gtt cag cca atg ctc atg aac agc ttc caa
Y T N M M E T V L D L S R R P S Q Q M P V Q P M L M N S F Q
969/301      999/311      1029/321
acc aac aag aac atc aaa cca gag gtg tat act acc agc cct cag ata cct gta tct acc act tca agc cag gtg agc ttg ttt gcc aac
T N K N I K P E V Y T T S P Q I P V S T T S S Q V S L F A N
1059/331     1089/341     1119/351
caa gag ccg tgt cac atg tca aca aca cag ggc gga acc tat gga caa atc tct cct att tca gat tct ggt gtc agt gac acc tcc cca
Q E P C R M S T T Q G G T Y G Q I S P I S D S G V S D T S P
1149/361     1179/371
gag cca agt tca gac tgg gaa gaa aat gtt gct tct gtg ctc ctc aac ctt taa
E P S S D W E E N V A S V L L N L *

```

NO PPVH motif

Equivalent region bracketed

Figure 13 (page 7 of 12)



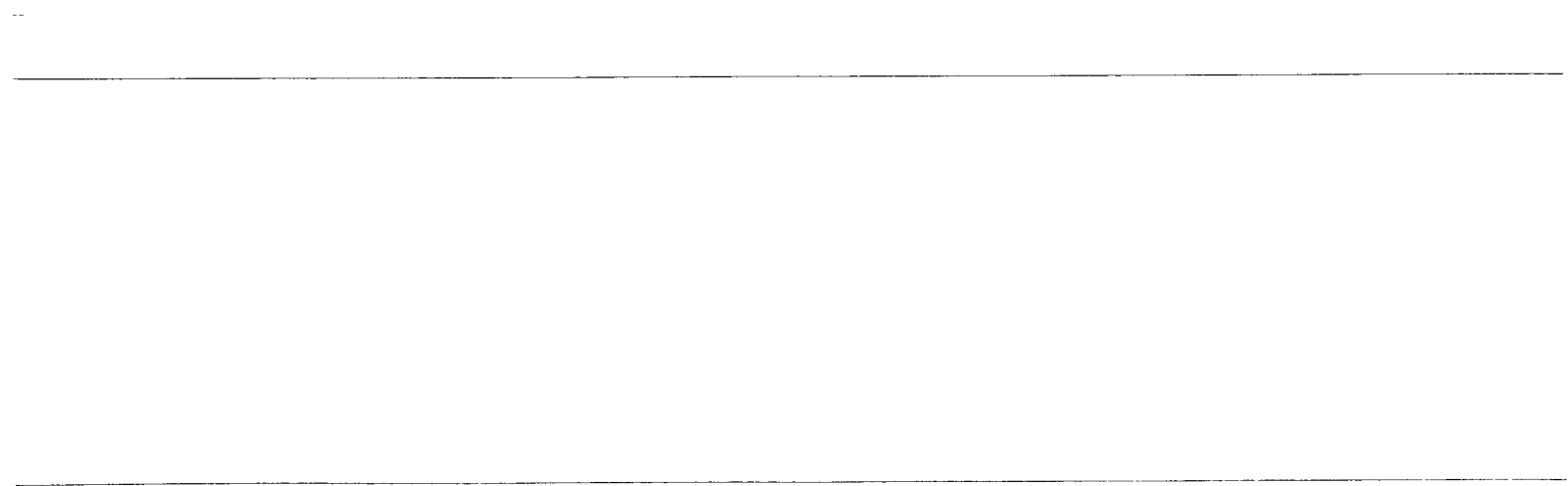
Mix.2 ORF [1 to 1110] -> Translate • 1-frame

DNA sequence 1110 bp ATGAATGGATTC ... CTGCATCTTTGA linear

```
1/1      31/11      61/21
ATG AAT GGA TTC AGC CAA CAA CTG GAG GAC TTC TAC OCT TCT TAC TTC TCC CCC AGC CCA TTG GGT TTC AGT GAG CCA GAG GTC CAG CCA
M N G F S Q Q L E D F Y P S Y F S P S P L G F S E P E V Q P
91/31      121/41      151/51
GTG GCC ATG AAC TTG GTG CCA ACA ATA CAG AAG GAC ATC CAG CAG CAG CCC AAC AGG AAA GAA GTC ACC AAG ATC CCC AGA GCC AGC GAG
V A M N L V P T I Q K D I Q Q Q Q P N R K E V T K I P R A S E
181/61      211/71      241/81
CAG AGC CCA GTC CAA AAT GTC AGG CCC AAA GAG GCC ATT AAT ACC AAA GAG GCT GAC TCG AGA AAC CCT GAG CCC GAC TCC TCT TTG GTT
Q S P V Q N V R P K E A I N T K E A D S R N P E P D S S L V
271/91      301/101      331/111
TCA GCA TCT CAG CGA CGG AAA AGG ACC TTC TTT ACC CAG GCC CAG CTG GAT ATC CTA GAG CAG TTC TTC CAA ACA AAC ATG TAC CCA GAC
S A S Q R R K R T F F T Q A Q L D I L E Q F F Q T N M Y P D
361/121      391/131      421/141
ATC CAC CAC CGG GAG GAG CTA GCG AGG CAC ATT TAC ATC CCT GAG TCC CGT ATT CAG GTC TGG TTC CAG AAC AGA AGA GCA AAG GTC AGA
I H H R E E L A R H I Y I P E S R I Q V W F Q N R R A K V R
451/151      481/161      511/171
CGT CAA GGT GCC AAA GCC ACC AAG CCC GCT CTT GCA AGC CAT CAT TAT TCT AGC ACC TCT GGG GCA ATG TTT CCT TCA GCA CCT GCC CCT
R Q G A K A T K P A L A S H H Y S S T S G A M F P S A P A P
541/181      571/191      601/201
AAC AGC TCC TCA TAC CAA ATG ACT TCA TCC CGG GCA CAG GTC CAG CCT CCA AAA GAA TAC CAA CTG AAT AAG TTT CAC CAA AGC CAG GGA
N S S S Y Q M T S S R A Q V Q P P K E Y Q L N K F H Q S Q G
661/221      691/231      721/241
CTT TCC TAC CCA GAC TCA TCT AGT GAT GTC TCG AGG CAG AGG TTC CTC TTG TCA CAG GCA ACA CCT GGT GTC TAC CAC CTG CCC CAG
F L S Y P D S S S D V S R Q R F L L S Q A T P G V Y H L P Q
751/251      781/261      811/271
GCA TCA TCC AAT GTC TAC GAT CAG AAT GTG AAA TCA AAT GAC CCT CTG TGG GGC CAG CAG CAA GTA TAT ACA AAT ATG GAG TCG GTC TTG
A S S N V Y D Q N V K S N D P L W G Q Q Q V Y T N M E S V L
841/281      871/291      901/301
AAC CTC AGC AGA AGA CCC CAG CAG ATG CCA GCT CAG CCA ATG TTC ATG AAC AGC TTC CAG ACC AAC AAA ATC ATC AAA TCA AAG ATG GAT
N L S R R P Q Q M P A A Q P M F M N S F Q T N K I I K S K M D
931/311      961/321      991/331
ACT ACC AGC CCT CCG ATC CCT GTA TCT ACC ACT TCA AGC CAC CAC AGT CAG ATG AGT TTG TTT GCC GGC CAA GAT CCA TGT CAC ATG TCA
T T S P P I P V S T T S S H H S Q M S L F A G Q D P C H M S
1021/341      1051/351
ACA GCA CCG GGC GGA ACC TAT GGA CAG ATT TCC CCC ATC TCA GAT TCT GGT GTC AGT GAC ACC TCC CCA GAG CCA AGT TCA GAC TGG GAA
T A P G G T Y G Q I S P I S D S G V S D T S P E P S S D W E
1081/361
GAG AAT GTT TCT GTG CTC CTG CAT CTT TGA
E N V S V L L H L *
```

Mix2 no PPRU motif.  
equiv. region bracketed.

Figure 13 (page 2 of 12)



Fast1 ORF [1 to 1557] -> Translate • 1-frame

*Xenopus*

DNA sequence 1557 bp atgggagacccc ... ggccttatgtag linear

*Xenopus*

1/1	31/11	61/21
atg gga gac ccc tcc agt ctg tac tca gga	ttc cca gct gga tcc cag tat gaa agt gtg	gag cct ccc agc ctt gcc ctg ctg agc tct
M G D P S S L Y S G	F P A G S Q Y E S V	E P P S L A L L S S
91/31	121/41	151/51
ata gac cag gag cag ctc cca gtg gcc acc	ggc cag tcc tat aat cac agt gtc cag cct	tgg ccc caa cct tgg cca ccc ttg tcc ctg
I D Q E Q L P V A T	G Q S Y N H S V Q P	W P Q P W P P L S L
181/61	211/71	241/81
tac aga gag ggg ggc acg tgg agc cca gac	aga ggc agt atg tat gga ctc tcc ccc ggc	acc cac gag ggc tcc tgc acc cac act cac
Y R E G G T W S P D	R G S M Y G L S P G	T H E G S C T H T H
271/91	301/101	331/111
gag ggc ccc aag gac tca atg gca gga gac	cat acc agg tcc agg aag agc aaa aag aag	aat tat cat cga tat tac aag ccc ccc tat
E G P K D S M A G D	H T R S R K S K K K	N Y H R Y Y K P P Y
361/121	391/131	421/141
tcc tac ctg gct atg att gcc ctg gtc atc	cag aac tcg ccc gag aag agg ctc aaa ctc	tcc cag atc ctg aag gag gtc agt aca ctc
S Y L A M I A L V I	Q N S P E K R L K L	S Q I L K E V S T L
451/151	481/161	511/171
ttc ccc ttc ttt aat ggg gat tat atg ggt	tgg aaa gac tcc atc agg cac aac ttg tct	tcc agt gac tgc ttt aag aag att ctc aaa
F P F F N G D Y M G	W K D S I R H N L S	S S D C F K K I L K
541/181	571/191	601/201
gac cct gga aag ccc cag gcc aag ggt aac	ttc tgg acg gtg gat gtt agc cgg att cct	ctg gat gcg atg aag ctg cag aac act gcg
D P G K P Q A K G N	F W T V D V S R I P	L D A M K L Q N T A
61/211	661/221	691/231
acc cga ggt gga tca gac tac ttt gtc	cag gat ttg gct cca tac atc cta cat aac	tat aaa tat gag cac aat gca ggg gcg tat
L T R G G S D Y F V	Q D L A P Y I L H N	Y K Y E H N A G A Y
721/241	751/251	781/261
ggg cac cag atg cct cca agt cat gcc aga	tcc ctg tct ttg gca gag gac tct caa cag	acc aac act ggt ggc aaa ctt aac aca tcc
G H Q M P P S H A R	S L S L A E D S Q Q	T N T G G K L N T S
811/271	841/281	871/291
ttt atg att gat tcc cta ctc cat gac ctg	caa gag gtg gat ctg cct gat gcc tcc agg	aac ctt gag aac caa agg atc tct ccg gct
F M I D S L L H D L	Q E V D L P D A S R	N L E N Q R I S P A
901/301	931/311	961/321
gta gcc atg aac aat atg tgg agc tct gct	cct ctt ctc tac act cat tcc aag cca aca	agg aat gcc aga agc cct ggt ttg tcc acc
V A M N N M W S S A	P L L Y T H S K P T	R N A R S P G L S T
991/331	1021/341	1051/351
atc cat tcc acg tac tcc tct tcc agc tcc	agc att tct aca atc tcc ccc gtt ggg ttt	cag aag gag cag gag aaa agt ggt cga caa
I H S T Y S S S S S	S I S T I S P V G F	Q K E Q E K S G R Q
1081/361	1111/371	1141/381
act caa agg gtt ggc cat ccc att aaa cga	tca aga gag gac gat gac tgc agt acc aca	tct tca gat cct gac act ggg aac tac tct
T Q R V G H P I K R	S R E D D D C S T T	S S D P D T G N Y S
1171/391	1201/401	1231/411
ccc att gag ccc cca aag aag atg ccc ttg	ctt tca ttg gac ttg ccc act tct tac aca	aag agt gtg gca cct aat gta gtg gca cca
P I E P P K K M P L	L S L D L P T S Y T	K S V A P N V V A P
1261/421	1291/431	1321/441
cca agt gtc ctg ccc ttc ttt cat ttt cct	cgc ttc acc tac tat aat tat gga cct tca	ccc tac atg acc cca cca tac tgg ggt ttt
P S V L P F F H F P	R F T Y Y N Y G P S	P Y M T P P Y W G F
1351/451	1381/461	1411/471
cca cat cct aca aat tct ggt ggg gat agt	cca cgt gga ccc caa tct cct ctg gac cta	gac aac atg tta cgg gcc atg cca ccc aac
P H P T N S G G D S	P R G P Q S P L D L	D N M L R A M P P N
1441/481	1471/491	1501/501
ag agt gtg ttt gat gtg ttg aca agt cac	cca ggt gac ctc gtc cat ccg tcc ttc ctc	agt caa tgc ttg ggc agc agt ggt tcc ccg
S V F D V L T S H	P G D L V H P S F L	S Q C L G S S G S P
1531/511		
tac cca agc aga caa ggc ctt atg tag		
Y P S R Q G L M *		

470-493 of *Xenopus* Fast1.

equivalent to 25aa region of Mixel.

Figure 13 (page 9 of 12)



HS FastIP.str [580 to 1677] -> Translate \* 1-frame

Human

DNA sequence 1793 bp gttgagtcgaatg ... cagagcttggtg linear

Vogelstein HS FastIP

human

```

580/1          610/11          640/21
atg ggg ccc tgc agc ggc tcc cgc ctg ggg ccc cca gag gca gag tgc ccc tcc cag ccc cct aag agg agg aag aag agg tac ctg cga
M G P C S G S R L G P P E A E S P S Q P P K R R K K R Y L R
670/31          700/41          730/51
cat gac aag ccc ccc tac acc tac ttg gcc atg atc gcc ttg gtg att cag gcc gct ccc tcc cgc aga ctg aag ctg gcc cag atc atc
H D K P P Y T Y L A M I A L V I Q A A P S R R L K L A Q I I
760/61          790/71          820/81
cgt cag gtc cag gcc gtg ttc ccc ttc ttc agg gaa gac tac gag ggc tgg aaa gac tcc att cgc cac aac ctt tcc tcc aac cga tgc
R Q V Q A V F F F R E D Y E G W K D S I R H N L S S N R C
850/91          880/101          910/111
ttc cgc aag gtg ccc aag gac cct gca aag ccc cag gcc aag ggc aac ttc tgg gcg gtc gac gtg agc ctg atc cca gct gag gcg ctc
F R K V P K D P A K P Q A K G N F W A V D V S L I P A E A L
940/121          970/131          1000/141
cgg ctg cag aac acc gcc ctg tgc cgg cgc tgg cag aac gga ggt gcg cgt gga gcc ttc gcc aag gac ctg ggc ccc tac gtg ctg cac
R L Q N T A L C R R W Q N G G A R G A F A K D L G P Y V L H
1030/151          1060/161          1090/171
ggc cgg cca tac cgg cgg ccc agt ccc cgg cca cca ccc agt gag ggc ttc agc atc aag tcc ctg cta gga ggg tcc ggg gag ggg gca
G R P Y R P P S P P P P P S E G F S I K S L L G G S G E G A
1120/181          1150/191          1180/201
tgg cgg ggg cta gct cca cag agc agc cca gtt cct gca ggc aca ggg aac agt ggg gag gag gcg gtg ccc acc cca ccc ctt ccc
W P G L A P Q S S P V P A G T G N S G E E A V P T P P L P
1210/211          1240/221          1270/231
tct tct gag agg cct ctg tgg ccc ctc tgc ccc ctt cct ggc ccc acg aga gtg gag ggg gag act gtg cag ggg gga gcc atc ggg ccc
S S E R P L W P L C P L P G P T R V E G E T V Q G G A I G P
1300/241          1330/251          1360/261
tca acc ctc tcc cca gag cct agg gcc tgg cct ctc cac tta ctg cag ggc acc gca gtt cct ggg gga cgg tcc agc ggg gga cac agg
S T L S P E P R A W P L H L L Q G T A V P G G R S S G G H R
1390/271          1420/281          1450/291
gcc tcc ctc tgg ggg cag ctg ccc acc tcc tac ttg cct atc tac act ccc aat gtg gta atg ccc ttg gca cca cca ccc acc tcc tgt
A S L W G Q L P T S Y L P I Y T P N V V M P L A P P P T S C
1480/301          1510/311          1540/321
ccc cag tgt ccg tca acc agc cct gcc tac tgg ggg gtg gcc cct gaa acc cga ggg ccc cca ggg ctg ctc tgc gat cta gac gcc ctc
P Q C P S T S P A Y W G V A P E T R G P P G L L C D L D A L
1570/331          1600/341          1630/351
ttc caa ggg gtg cca ccc aac aaa agc atc tac gac gtt tgg gtc agc cac cct cgg gac ctg gcg gcc cct ggc cca ggc tgg ctg ctc
F Q G V P P N K S I Y D V W V S H P R D L A A P G P G W L L
1660/361
tcc tgg tgc agc ctg tga
S W C S L *

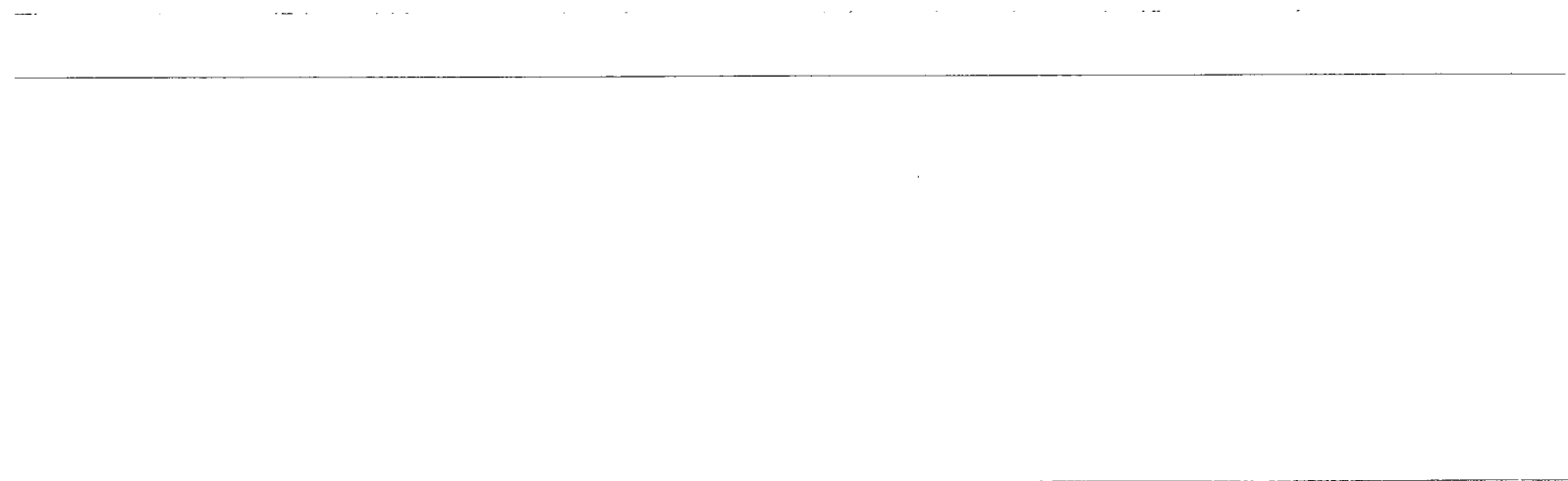
```

bracketed 25 amino acids equivalent to Mixer region

24 dx in this case

327-350

Figure 13 (page 10 of 12)





Fast2 ORF [1 to 1206] -> Translate • 1-frame

Mouse

DNA sequence 1288 bp atggcctcgggc ... gcatgtccgcag linear

Mouse

```

1/1          31/11          61/21
atg gcc tgg ggc tgg gac ctg gcc tca act tac act ccg act acc ccg agc ccc cag tta gcc ctg gct ccg gcc cag ggc tac ctg cct
M A S G W D L A S T Y T P T T P S P Q L A L A P A Q G Y L P
91/31          121/41          151/51
tgt atg ggg cct cgc gac aac tct cag ctg agg cct cca gag gca gag tct ctt tgg aag acc ccc aag agg agg aag aga tac cta
C M G P R D N S Q L R P P E A E S L S K T P K R R K K R Y L
181/61          211/71          241/81
cgg cat gac aag ccc ccc tac acc tac ttg gcc atg atc gcc ttg gta att cag gcc gca ccc ttc cgc agg ctg aaa ctg cct cag att
R H D K P P Y T Y L A M I A L V I Q A A P F R R L K L A Q I
271/91          301/101          331/111
atc cgt cag gtc cag gca gtg ttc ccc ttc ttg agg gac gac tat gag ggc tgg aag gac tcc atc cgc cac aac ctt tcc ttt aat cgg
I R Q V Q A V F P F F R D D Y E G W K D S I R H N L S S N R
361/121          391/131          421/141
tgc ttc cat aag gtg ccc aag gac cct gca aag ccc cag gcc aag ggc aac ttc tgg gcg gtg gat gtg agc ctg att cct cct gag gcg
C F H K V P K D P A K P Q A K G N F W A V D V S L I P A E A
451/151          481/161          511/171
ctg cgc ctt cag aac act gcc ctg tgc cgt cga tgg cag aac cgg ggc acc cac aga gct ttc gcc aag gac ctg agc ccc tac gtg ctg
L R L Q N T A L C R R W Q N R G T H R A F A K D L S P Y V L
541/181          571/191          601/201
cac ggc cag cct tat cag cca ccc agt ccc cca cca cca cct agg gag ggt ttc agc atc aag tcc ctg cta ggg gac ctt ggc aaa gaa
H G Q P Y Q P P S P P P P P R E G F S I K S L L G D L G K E
631/211          661/221          691/231
c aca tgg ccc aag cat cct ggg ctg ctt gga cag agc act gca gct cag gca ggc acc ttg tca aag ggg gaa gaa ggg atg ggc act
S T W P K H P G L L G Q S T A A Q A G T L S K G E E G M G T
721/241          751/251          781/261
gga ccc tct agc tcc tct gag acg cct ctg tgg ccc ctg tgc tcc ctt cct ggg ccc aca atc ata gag ggg gag agt tcc caa ggg gag
G P S S S S E T P L W P L C S L P G P T I I E G E S S Q G E
811/271          841/281          871/291
gta atc agg cct tct ccc gtc acc cca gat caa ggc tcc tgg ccc ctg cac tta ctt gag gat tcc gca gat tcc agg gga gtg ccc agg
V I R P S P V T P D Q G S W P L H L L E D S A D S R G V P R
901/301          931/311          961/321
agg ggg agc aga gcc tcc ttg tgg gga cag cta ccc act tct tac ttg ccc atc tat acg ccc aat gta gta atg ccc ttg gcc aca cta
R G S R A S L W G Q L P T S Y L P I Y T P N V V M P L A T L
991/331          1021/341          1051/351
ccg acc acc tct tgt ccc cag tgc cca tct tct gcc agc cca gct tac tgg agc gta ggc act gaa tcc caa ggg tcc cag gac ctg ctg
P T T S C P Q C P S S A S P A Y W S V G T E S Q G S Q D L L
1081/361          1111/371          1141/381
tgt gat cta gac tcc ctg ttc cag gga gta cca ccc aac aag agt atc tat gat gtg tgg gtc agc cat cct agg gac ctg gca gct cct
C D L D S L F Q G V P P N K S I Y D V W V S H P R D L A A P
1171/391          1201/401
gcc cca ggc tgg ctg ctt tcc tgg tac agc atg taa
A P G W L L S W Y S M

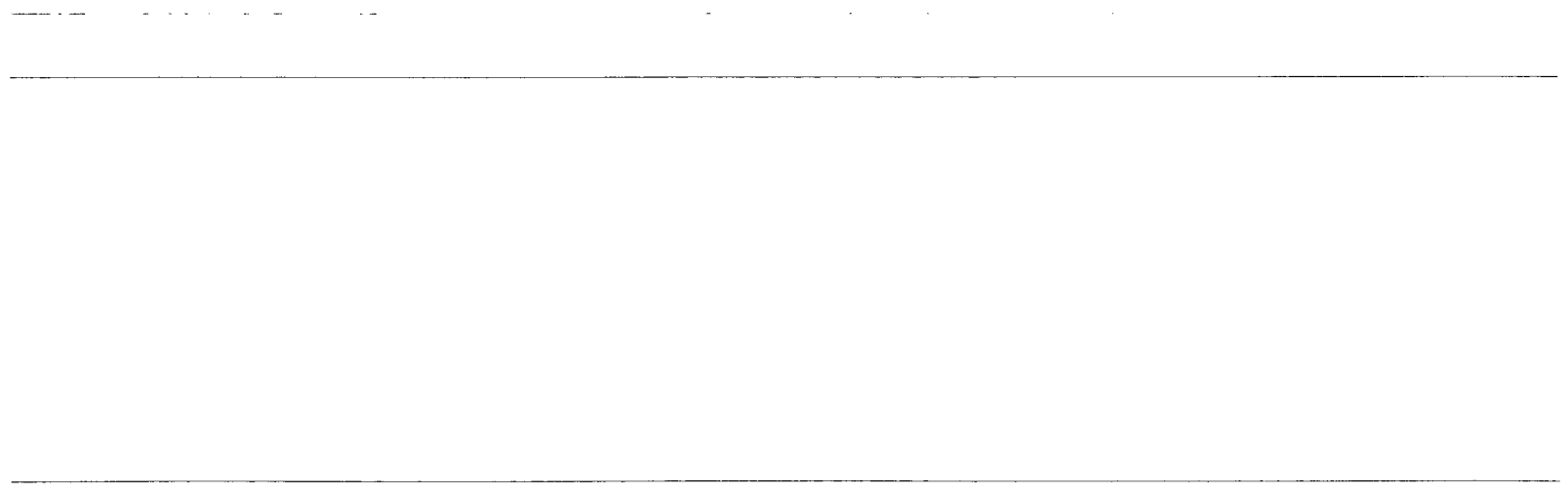
```

mouse fast 1.

363-386.

equiv. + 25 x1 region Mixer

Figure 13 (page 11 of 12)



# Xenopus FAST 3

Protein sequence 339 aa VAAALELVDPG ... DPIAYSEGLQLA

	10	20	30	40	50	60	70	80	90	100	
1	VAAALEL	VDP	PCRNSARDA	KPPYSYLAMI	SLVTQNSPEK	RLKLSQILQD	ISSLFPPFFKG	NYOGWKDSIR	HNLSSNDCFR	KVLKDPLKPO	AKGNYWTVDV 100
101	TRIPPDALKL	QNTAVTRQDL	FPLDLAPYIL	HQOPYRSLER	LSANHTRGRT	TPRMEPEVQI	PVSDPAVSFP	HILWNLPTSY	SKCVAPNVVA	PPSIHPILLY	200
201	SNFPSISIYN	YLPPPYGSPV	YSDRRELLAF	GLHPQIPLTP	KPEELQNAFS	DFFPNKTVFD	IPVYTGHGPF	LASQSLFSPH	LPTGTPLWA	TGRLGVEGPI	300
301	*LYLWICKYC	HKYIYI*KKK	KKKEGGDPI	AYSEGLQLA							339
	10	20	30	40	50	60	70	80	90	100	

XFast-3 seq around PPNu motif.

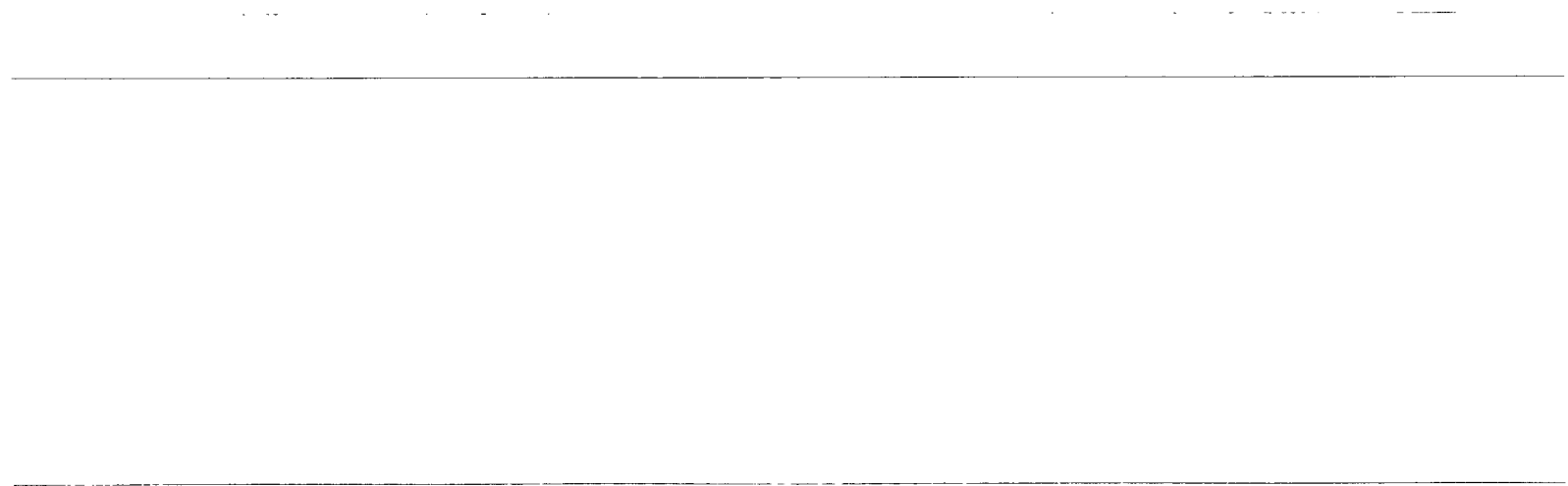
partial C-terminal  
coding sequence

FPPNKTVFDIPVYTGHGPF	fast 3
MPPNKSVFDVLTSHPGD	fast 1

align better

FPPNKTVFDIPVYTGHGPF	fast 3
MPPNKSVFDV--LTSHPGD	fast 1
VPPNKSIYDV--WVSHPRD	new fast 2

Figure 13 (page 12 of 12)



Chicken Mix (210 AA)

MAALRFGPPPAELPAVPPSCPPGRWLCGTAGGSGGGPGAAPAPLASLPPA  
AEGAPSAQRRKRTSFTAAQLETLELVFQDTMYPDIYLRERLADATQIPES  
RIQVWFQNRRAKSRRQRGPPRPGAPAPPPPPPQRSPCGAAPLLRAREEHR  
EWPPRAAGPPGSALRPHGGSGGAPAGPYPPRPAFPLPAGGGFSELGTEWE  
ENAIGAFRAL

Figure 14 (page 1 of 2)

1

---

---

Mix.1 prot x cmix.prot=> Protein Alignment

Protein sequence 1 377 aa MDGFSQLEDLY ... WEENVASVLLNL  
 Protein sequence 2 210 aa MAALRFGPPPAE ... WEENAIGAFRAE

Layout: Standard  
 Method: Blocks (Martinez)  
 Block Length: 6-aa  
 Mismatch penalty: Smaller (1)  
 Gap penalty: Medium (2)  
 Weighting: BLOSUM62

```

      20      40      60      80     100
1 MDGFSQLEDLYPSCFSPCLGFEPVIOQFAMNLAPAAQKDFQOHFNRRVTKIPGAGEQSPVQNVVRPKDAINPKADPRSP
  AA + V G A S A P+QRRKR 100
=====MAALRFGPPPAELPAVPPSPGRLCGTAGGSGGGPGAAPPLASLPPAAEGAPSAQRRKR 62
      120      140      160      180     200
101 TFFTQAQLDILEOFFQTNMYPDIIHREELARHIYIPESRIQVWFQNRRAKVRROGAKATKPILAGHHYSGTSGANRAMF
  T FT AOL+ LE FO MYPD1+ RE LA 1PESRIQVWFQNRRAK RRO A P A H+ RA
63 TSFTAAQLETLELVFQDITMYPDIYLRERLADATQIPESRIQVWFQNRRAKSRROGPPRPGAPAPPPPPQSPCGAA=
  80 100 120 140 160 180 200
=====PLLRAREE==HREWPPRAAGP 159
      220      240      260      280     300
201 PLKESQNMFMHONQGFPLPYDSSCNVSRQFLMSQPTPGAYHLPOASSNVYNQNVKSONPLWNQQTINRVYTNNMETV
  P + + PYP S+ +
160 PGSALRPHGGSGGAPAGPYPRPAFPLPAGGGFSELGTWEENAIGAFRAL===== 210
      160      180      200      220      240      260      280      300
301 TNKNIKPEVYTTSPQIPVSTTSSQVSLFANQEPCHMSTTQGGTYGQISPISDSGVSDTSPSPSSDWEENVASVLLNL
  160 180 200 220 240 260 280 300
===== 377
    
```

% Identity = 16.7 (63/377)

Figure 14 (page 2 of 2)

PCT 2800 65

4-9-00

Eric Potter clarkson